SOME ASPECTS OF INFORMATION THEORY OF ORDER α

A THESIS

Submitted in Partial Fulfilment of the Requirements for

THE DEGREE OF DOCTOR OF PHILOSOPHY

By

P. N. Chhabra, *M. A. (Math.), M. A. (Math.Stat.)*

to the
Department of Mathematics
INDIAN INSTITUTE OF TECHNOLOGY
KANPUR

October, 1968

In the unforgettable

memory of

My Dearest Mother

C E R T I F I C A T E

This is to certify that the thesis entitled SOME ASPECTS OF INFORMATION THEORY OF ORDER α by P. N. Chhabra, for the award of the Degree of Doctor of Philosophy, I.I.T. Kanpur, is a record of bonafide research work carried out by him under my supervision and guidance for the last three years. The thesis has, in my opinion, reached the standard fulfilling the requirements for the Doctor of Philosophy degree. The results embodied in this thesis have not been submitted to any other university or institute for the award of any degree or diploma.

20th October, 1968.

(J. N. KAPUR)
PROFESSOR AND HEAD,
DEPARTMENT OF MATHEMATICS.

A C K N O W L E D G E M E N T S

I feel extremely indebted to my supervisor Prof. J.N. Kapur, Ph.D., F.N.A.Sc., F.A.Sc., F.I.M.A.(U.K.), Professor and Head of the Department of Mathematics, I. I. T., Kanpur, for suggesting to me the problems and for his great encouragement and ever-readiness to discuss them with me. It is again to Prof. Kapur that I owe a deep sense of appreciation for his constant watch over my progress and opportune help in removing the pitfalls I encountered in the writing of my thesis.

I am also grateful to Prof. R.S. Varma, Dean of Faculty of Mathematics, University of Delhi, whose inspiring seminars in 1964 first introduced me to the subject of information theory.

I deem it my duty to express my sincere thanks to Dr. P.K. Kelkar, Director, I.I.T.,Kanpur for awarding me the much-needed scholarship.

I would also very much like to give my thanks to the faculty members of the department of mathematics responsible for offering the courses I credited as a part of my Ph. D. program. I gained a lot from their inspiring teaching. I am also thankful to all other faculty and staff members of the department for their kind cooperation and for the help they have rendered to me in my work. I am also thankful to Mr. S. S. Pethkar for typing the thesis so nicely.

TABLE OF CONTENTS

CHAPTER 3

DISCRETE MEMORYLESS CHANNELS

CHAPTER 4

SHANNON'S FUNDAMENTAL CODING THEOREMS

CHAPTER 5

α- ENTROPIES FOR CONTINUOUS MEMORYLESS CHANNELS

PART I

PART II

CHAPTER 6

FURTHER ASPECTS OF MULTIDIMENSIONAL CONTINUOUS MEMORYLESS CHANNELS

CHAPTER 7

SEMI-CONTINUOUS MULTIDIMENSIONAL MEMORYLESS CHANNELS

CHAPTER 8

STATIONARY MARKOV CHAINS AND STATIONARY SYMMETRIC CHANNELS

PART I

PART II

LIST OF MAIN SYMBOLS

$H_\alpha(X)_g$	:	Renyi's α-entropy associated with a generalised discrete probability scheme.
$H_\alpha(X), H_\alpha(Y), H_\alpha(X,Y)$ $H_\alpha(Y/X)$, $H_\alpha(X/Y)$	:	α-entropies associated with two univariate probability schemes specified by the random variables X and Y or equivalently α-entropies associated with a two-dimensional channel.
$I_\alpha(X ; Y)$	:	α-information conveyed about X by Y or equivalently α-information of type I.
$I_\alpha(Y ; X)$	:	α-information conveyed about Y by X or equivalently α-information of type II.
$C_\alpha^{(1)}$	:	α-capacity of type I .
$C_\alpha^{(2)}$	:	α-capacity of type II.
$I_\alpha(X; Y_1, Y_2, \ldots, Y_m)$	:	α-information conveyed about X when an input is received by m outputs.
$I_\alpha(X; Y_1; Y_2)$	:	α-information for three alphabets X, Y_1 and Y_2.

The symbols with the subscript α replaced by 1, e.g. $H_1(X)$, $I_1(X;Y)$ etc., will represent the corresponding Shannon's definitions.

All other symbols used are understood from the text.

Prefix 'α' stands for ''of order α'', e.g. 'α-entropy' would mean 'entropy of order α'.

The unspecified base of the logs is assumed to be 2 but wherever a change of base is needed for operational facilities in differentiation and integration, such a change has been indicated in the start e.g. from section 2 onwards

of part I, chapter 5,base of the logs has been assumed to be e and this change has been indicated at the end of section 1, part I, chapter 5. Very occasionally the base is even mentioned just for the sake of emphasis.

SYNOPSIS

SOME ASPECTS OF INFORMATION THEORY OF ORDER α.

A thesis submitted
In partial fulfilment of the requirements
for the Degree of

DOCTOR OF PHILOSOPHY

by

P. N. Chhabra, M.A.(Math.), M.A.(Math.Stat.)

to the

Department of Mathematics,
Indian Institute of Technology, Kanpur.

October 1968.

In the present thesis, taking Renyi's α-entropy (entropy of order α, $\alpha > 0$, $\alpha \neq 1$) discussed in his paper ''On Measures of Entropy and Information, Proc.4th Berkeley Symp. Math. Stat. Prob., 1961, pp. 547-561'' as the basic measure of uncertainty, other α-entropies $\alpha > 0$, $\alpha \neq 1$ associated with a channel are defined and efforts have been made to generalise most of the basic concepts available in the present-day information theory. All the results, except the two results on maximisation of $\bar{H}_\alpha(\bar{X})$ and the evaluations of α-capacities based on one of them, developed in the thesis are true for $\alpha > 0$, $\alpha \neq 1$ and tend in the limit, as $\alpha \to 1$, to the corresponding available results. The whole of the thesis, except the section on 'Generalised Entropies' in chapter 1 on Introduction, deals with complete distributions.

Chapter 1 introduces the concept of generalised entropies and gives the chapter-cum-section-wise summary of other chapters.

In Chapter-2, α-entropies $\alpha > 0$, $\alpha \neq 1$, associated with two discrete probability schemes specified by the random variables X and Y are defined as follows:

$$H_\alpha(X) = (1/1-\alpha) \log \sum_i p^\alpha(x_i)$$

$$H_\alpha(Y) = (1/1-\alpha) \log \sum_j p^\alpha(y_j)$$

$$H_\alpha(X,Y) = (1/1-\alpha) \log \sum_{ij} p^\alpha(x_i, y_j)$$

$$H_\alpha(Y/X) = (1/1-\alpha) \sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i)$$

$$H_\alpha(X/Y) = (1/1-\alpha) \sum_j p(y_j) \log \sum_i p^\alpha(x_i/y_j)$$

Some of the properties of these α-entropies are discussed. Some basic inequalities connecting them and tending in the limit, as $\alpha \to 1$, to Shannon's basic results of the types

$$H_1(Y/X) \leq H_1(Y)$$

and $$H_1(X,Y) = H_1(X) + H_1(Y/X)$$

are developed. All the results developed for two discrete probability schemes are extended for any finite number of such schemes. In the last section, alternative definitions of conditional α-entropies

$$H_\alpha(Y/X) = (1/1-\alpha) \log \sum_i p(x_i) \sum_j p^\alpha(y_j/x_i)$$

$$H_\alpha(X/Y) = (1/1-\alpha) \log \sum_j p(y_j) \sum_i p^\alpha(x_i/y_j)$$

are developed and the truth of various results developed in the chapter is ascertained in the light of these new definitions.

In Chapter 3, two types of α-informations, $\alpha > 0$, $\alpha \neq 1$, tending in the limit, as $\alpha \to 1$, to the available Shannon's concept of mutual information, are developed and based on them, the concepts of α-capacities, α-redundancies and α-efficiencies are defined. In the light of these concepts of order α the characteristics of noiseless channels, channels with independent input and output, channels with symmetric noise structure, BSC, BEC and cascaded channels are considered. α-Information for a single input and several outputs is also defined and the available additive property of corresponding mutual information is generalised i.e. proved valid for the concept of α-information. α-information for more than two alphabets is also defined. All these results are ascertained to be valid for alternative definition of $H_\alpha(Y/X)$ and $H_\alpha(X/Y)$.

In Chapter 4 a generalisation, based on the concepts of α-information of type II and α-capacity of type II, of Shannon's second fundamental theorem for noisy discrete memoryless channels is stated and an heuristic proof for BSC is given. Available Fano bound for discrete

memoryless channels is generalised and as a consequence of this generalisation, a generalised form of the converse of Shannon's second fundamental theorem is stated and the proof is given for a family of extensions of a discrete memoryless channel. A generalisation, based on $H_\alpha(X/Y)$, of Shannon's theorem for noiseless coding is also made. α-Bounds for $\bar{L} = \sum_i p(x_i)\, n_i$ in terms of $H_\alpha(X)$ are developed and two results, one dealing with Shannon-Fano encoding and the other known result on the attainment of $\bar{L}$ its lower bound, are derived as corollaries of these α-bounds. Some problems that arise from these bounds are also stated. All the results developed in the chapter are ascertained to be valid for alternative definitions of conditional α-entropies.

In Chapter 5 the notion of Renyi's α-entropy is extended for continuous and multidimensional continuous schemes. Based on these extensions, other α-entropies associated with a bivariate continuous and a multidimensional continuous channel are defined and the various results developed in Chapter 2 for discrete channels are extended for continuous and multidimensional continuous channels. Some mathematical drawbacks in adopting these extensions as measures of uncertainty are brought out and the concepts of α-informations, tending in the limit, as $\alpha \to 1$, to the available concept of the rate of information processed in such channels, are defined. It has been shown that these concepts of α-informations don't have the salient drawbacks found in the extensions

In Chapter 6 α-capacities for multidimensional continuous channels are defined and as a need for their evaluations the problem of maximisation of $H_\alpha(\bar{X})$ is tackled for the following three types of contraints on $\bar{X}$:

(i) when the domain of $\bar{X}$ is bounded i.e. $a_1 < X_1 < b_1, \ldots\ldots, a_n < X_n < b_n$

(ii) when the components of $\bar{X}$ assume only non-negative values and have specified first moments.

(iii) when the components of $\bar{X}$ have specified standard deviations.

α-Informations, when both the input $\bar{X}$ and output $\bar{Y}$ are normally distributed, are evaluated. α-capacities are also evaluated when the channel is subject to

(i) additive Gaussian noise and each component of the input vector is distributed with mean zero and a specified standard deviation.

(ii) besides conditions in (i), input is also Gaussian.

A generalisation of Shannon-Hartley channel-capacity formula is also made.

In Chapter 7 α-entropies associated with a multidimensional semi-continuous memoryless channel are defined, available Fano bound for such channels is generalised and as a consequence of this generalisation, a generalised form of the converse of Shannon's second theorem for semi-continuous memoryless channels is stated and its proof for a family of extensions of such a channel is outlined.

In part I of Chapter 8, α-entropies associated with different steps of a stationary Markov chain, when its initial state probabilities are specified, are defined and two theorems connecting these α-entropies are given. Some of the results of these theorems tend in the limit, as $\alpha \to 1$, to the known result on the additivity of corresponding Shannon's entropies.

In part II of Chapter 8, α-entropy associated with a stationary source is obtained and its existence is established when the source is identified with a stationary symmetric channel. The concepts of the rates of α-informations for stationary symmetric channels are also developed. These rates tend in the limit, as $\alpha \to 1$, to the known rate of information processed in such channels.

CHAPTER I

INTRODUCTION

1. GENERALISED ENTROPIES

Shannon (1948) gave for a complete discrete probability scheme

$$X = \begin{pmatrix} x_1, \ldots\ldots, x_n \\ p(x_1), \ldots\ldots, p(x_n) \end{pmatrix}, p(x_i) \geqslant 0, \; i = 1 \text{ to } n, \sum_{i=1}^{n} p(x_i) = 1 \qquad \ldots(1)$$

the expression

$$H_1(X) = -\sum_i p(x_i) \log p(x_i) \qquad (2)$$

as a measure of entropy. The quantity $H_1(X)$ has been characterised by various sets of postulates ; a weaker set given by Fadeev (1957) is as follows :

(i) $H_1(X)$ is a symmetric function of $p(x_i)$

(ii) $H_1(p(x), 1-p(x))$ is a continuous function of $p(x)$ for $0 \leqslant p(x) \leqslant 1$

(iii) $H_1(1/2, 1/2) = 1$

(iv) $H_1(t\,p(x_1), (1-t)p(x_1), p(x_2), \ldots, p(x_n))$

$$= H_1(p(x_1), \ldots, p(x_n)) + p(x_1)\, H_1(t, 1-t)$$

for $0 \leqslant t \leqslant 1$

These postulates characterise $H_1(X)$ uniquely. The weakest set of postulates known at this time may be found in Lee (1964). Kapur (1967 b) has also made a study of these postulates.

The fact that $H_1(X)$ can be regarded as the mean value of the variable $\log_2 (1/p(x))$ was earlier emphasized by some anthors especially by Bernard (1951). So keeping (2) in view, Renyi (1961) defined the concept of a generalised discrete probability scheme

$$X = \begin{pmatrix} x_1, \ldots, x_r \\ p(x_1) \ldots, p(x_r) \end{pmatrix}, \; p(x_i) > 0, i = 1 \text{ to } r, 0 < \sum_{i=1}^{r} p(x_i) \leq 1 \qquad \ldots (3)$$

and defined the entropy attached to it as

$$H_1(X)_g = \frac{-\sum_{i=1}^{r} p(x_i) \log p(x_i)}{\sum_{i=1}^{r} p(x_i)} \qquad (4)$$

Renyi characterised $H_1(X)_g$ by the following five postulates:

1. $H_1(X)_g$ is a symmetric function of $p(x_i)$.

2. If $\{p(x)\}$ denotes the generalised probability distribution consisting of the single probability $p(x)$, then $H_1(\{p(x)\})_g$ is a continuous function of $p(x)$ for $0 < p(x) \leq 1$.

3. $H_1(\{1/2\})_g = 1$

The other two postulates need the introduction of some notations. Let Δ denote the set of all finite discrete generalised probability distributions associated with generalised discrete schemes. For $P(X), P(Y) \varepsilon \Delta$, let $P(X) * P(Y)$ denote the direct product of $P(X)$ and $P(Y)$ i.e. the distribution consisting of the numbers $p(x_i)p(y_j)$ with $i = 1, .., r$, $j = 1, \ldots, s$ and if $\sum_i p(x_i) + \sum_j p(y_j) \leq 1$ then let

$$P(X) \cup P(Y) = \{p(x_1), \ldots, p(x_r), p(y_1), \ldots, p(y_s)\} .$$

Note that if $\sum_i p(x_i) + \sum_j p(y_j) > 1$ then $P(X) \cup P(Y)$ is not defined. So the last two postulates are

4. $H_1(X * Y)_g = H_1(X)_g + H_1(Y)_g$

5. $H_1(X \cup Y)_g = \dfrac{(\sum_i p(x_i))\, H_1(X)_g + (\sum_j p(y_j)) H_1(Y)_g}{\sum_i p(x_i) + \sum_j p(y_j)}$

Further Renyi argued that if, in place of ordinary mean as used for getting (2) and laid down in postulate 5, we consider the Kolmogorov- Nagumo generalised mean of the random variable $\log_2 \frac{1}{p(x)}$ w.r.t. the strictly monotonic and continuous parametric function

$$f_\alpha(x) = 2^{(1-\alpha)x} \qquad \alpha > 0, \alpha \neq 1$$

or in other words if we replace the postulate 5 by postulate 5' :

$$H_1(X \cup Y)_g = f_\alpha^{-1}\left[\frac{\sum_i p(x_i) f_\alpha(H_1(X)_g) + \sum_j p(y_j) f_\alpha(H_1(Y)_g)}{\sum_i p(x_i) + \sum_j p(y_j)}\right]$$

we shall arrive at the concept of entropy of order α. For the generalised distribution given by (3), it is given by

$$H_\alpha(X)_g = \frac{1}{1-\alpha} \log_2 \frac{\sum_{i=1}^{r} p^\alpha(x_i)}{\sum_{i=1}^{r} p(x_i)} \qquad \alpha > 0, \alpha \neq 1 \qquad (5)$$

In particular when $\sum_i p(x_i) = 1$, (5) becomes

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2 \sum_i p^\alpha(x_i) \qquad \alpha > 0, \alpha \neq 1 \tag{6}$$

It may be noted that both $H_\alpha(X)_g$ and $H_\alpha(X)$ satisfy all the postulates 1,2,3,4 and 5$'$.

Varma (1965, 1966a) gave two generalisations of (5) by defining

$$H^n_{A^\alpha}(X)_g = \frac{1}{n-\alpha} \log \frac{\sum_i p^{\alpha-n+1}(x_i)}{\sum_i p(x_i)} \qquad (n-1) < \alpha < n, n \geqslant 1$$

$$H^n_{B^\alpha}(X)_g = \frac{n}{n-\alpha} \log \frac{\sum_i p^{\alpha/n} p(x_i)}{\sum_i p(x_i)} \qquad 0 < \alpha < n, n \geqslant 1$$

and studied their properties like monotonicity etc. It may be noted that these entropies are two parameter entropies. Later on Kapur (1967a) generalised (5) by defining generalised entropy of order α and type β :

$$H_\alpha^\beta(X)_g = \frac{1}{1-\alpha} \log \frac{\sum_i p^{\alpha+\beta-1}(x_i)}{\sum_i p(x_i)} \qquad \beta > 1, \alpha > 0, \alpha \neq 1 \quad \ldots \tag{7}$$

and studied its properties. It may be noted that in the limit as $\alpha \to 1$ and $\beta = 1$ (7) tends to (4) .

2. ABOUT THE PRESENT THESIS

In the present thesis taking (6) as the basic measure α entropy, other entropies of order α associated with a channel are defined and efforts have been made to generalise most of the concepts available in the present day information theory. It is possible that almost all the concepts involving Shannon's measure of information in the present day information theory may be generalised to give ''Information Theory of Order α'' in its full form. In that case the present available theory would be just a limiting case, as $\alpha \to 1$, of the generalised ''Information Theory Of Order α''. However the difficulties in developing this generalised theory in its full form are considerable.

The following is the chapter-wise summary of what is contained in this thesis:

Chapter - 2

In section 1 α- entropies associated with two discrete schemes are developed.

In sections 2.1 and 2.2 it is proved that these α- entropies are continuous and monotonically decreasing functions of α and in section 2.3 their maximality is discussed. In sections 2.4 and 2.5 an easier method for finding successive derivatives and bounds for second derivatives of these α- entropies are developed.

In section 3.3 Shannon's most basic and famous inequality $H_1(Y/X) < H_1(Y)$ is generalised i.e. proved for corresponding α-entropies. In section 3.5 two infinite sets of inequalities, each member of either set tending in the limit, as $\alpha \to 1$, to the basic Shannon's equality $H_1(X,Y) = H_1(X) + H_1(Y/X)$, and two bounds, one for $0 < \alpha < 1$ and the other for $\alpha > 1$, for $H_\alpha(X, Y)$ are developed. In section 3.4 Khinchin's generalisation of Shannon's inequality $H_1(Y/X) < H_1(Y)$ is generalised. In section 3.1 and 3.2 two more relations between these α-entropies are developed. In section 3.6 a study of the validity of $H_\alpha(X,Y) = H_\alpha(X) + H_\alpha(Y/X)$ for $\alpha > 0$ is made.

In section 4 various main results developed in section 3 are generalised for r discrete schemes.

In section 5 alternative definitions of conditional α-entropy and α-equivocation are given and the validity of various results developed in this chapter is explored in the light of these alternative definitions.

Chapter - 3

In section 1 α-entropies, associated with two schemes and defined in section 1 of chapter 2, are interpreted as those associated with a discrete memoryless channel.

In section 2 two types of α-informations and hence two types of α- capacities, α- redundancies and α- efficiencies are defined. Definition of two types of α-informations are the results of the invalidity in general of

$$H_\alpha(X,Y) = H_\alpha(X) + H_\alpha(Y/X) = H_\alpha(Y) + H_\alpha(X/Y) \qquad \alpha > 0,\ \alpha \neq 1$$

In section 3 α-informations, when corresponding to a single input a number of outputs are received, are defined and the available result

''Information conveyed about an input by a compound observation does not depend on whether we consider the compound observation as a whole or broken into its components'' is generalised.

In section 4 following McGill (1954) α-information of several alphabets is defined and expressed in terms of α- entropies.

In section 5, the available result on the leakage of information through a cascade of channels is generalised.

In section 6, characteristics of noiseless channels, channels with independent input and output, channels with symmetric noise structure, BSC and BEC are considered in the light of α- entropies and the concepts defined in section 2. Lastly in section 7 all these results are verified in the light of alternative definitions of conditional α-entropy and α- equivocation developed in section 5 of chapter 2.

Chapter - 4

In section 2 a generalisation of Shannon's second fundamental theorem for noisy discrete memoryless channels is stated by taking the information processed in the channel as given by α-information of type II and a heuristic proof is given for BSC.

In section 3 the available Fano bound of discrete channels is generalised and, as a consequence of this generalisation, a generalised form of the converse of Shannon's second fundamental theorem is stated and proved for a family of extensions of a discrete memoryless channels.

In Abramson (1963) is given a generalisation of Shannon's first fundamental theorem. According to this generalisation every received symbol makes us consider a D-ary code so as to find what has been communicated. The proof of this is based on the concept of equivocation. Campbell (1965) gave a generalisation of Shannon's first theorem for noiseless coding by considering an exponential cost function and taking Renyi's α-entropy as the basic measure of uncertainty. Following the lines of the generalisation given in Abramson (1963) and based on our concept of α-equivocation we have generalised Campbell's theorem and hence Shannon's theorem. This is done in section 4 .

In section 5.1, taking the usual average length $\bar{L} = \sum_i p(x_i)n_i$ associated with the transmission of a sequence of encoded words of lengths given by $\{n_i\}$ and respective probabilities of transmission given by $\{p(x_i)\}$, generalisations of the result $\bar{L} \geqslant \frac{H_1(X)}{\log D}$ are made. In section 5.2 on the basis of these generalised bounds of $\bar{L}$ two results, one stated on page 140, Reza (1961), on the Shannon-Fano encoding procedure and the other, a known result dealing with the necessary and sufficient condition on the attainment of $\bar{L}$ its lower bound, when the corresponding entropy is measured

by Shannon's definition, are discussed. Lastly in section 5.3 some problems that result from these bounds are given and a note is added as to the validity of all these results for the alternative definitions of conditional α-entropy and α- equivocation.

Chapter -5

Part I

In section 1 Renyi's α-entropy is extended for a continuous scheme and some of its requisite properties are stated.

In section 2 α-entropies associated with a two-dimensional continuous memoryless channel are defined.

In section 3 some relations between these α-entropies are developed. These relations run parallel to the corresponding relations developed for discrete channels in chapter 2. A salient feature of these relations is the validity of

$$H_\alpha (X,Y) = H_\alpha(X) + H_\alpha(Y/X) = H_\alpha(Y) + H_\alpha(X/Y) \quad \alpha > 0$$

when X and Y are jointly Gaussian.

In section 4 it is shown that $H_\alpha(X)$ for a continuous scheme, in comparison to $H_\alpha(X)$ for a discrete scheme, is not always positive, finite and invariant even under linear transformations.

In section 5 as in the discrete case two types of α-informations are defined and some of its properties are considered.

In section 6 α- capacities are defined. The procedure of defining α-capacities for continuous channels runs parallel to that of discrete channels.

In section 7 all the results developed in this part are considered in the light of alternative definitions of conditional α-entropy and α-equivocation.

<u>Part II</u>

In section 1 Renyi's α-entropy is extended to a multidimensional continuous scheme.

In section 2 α-entropies associated with a multidimensional continuous memoryless channel are defined and some basic relations between them are developed. These relations run parallel to the corresponding relations developed in part I of this chapter.

In section 3 all the relations developed in section 2 are generalised for r multidimensional continuous schemes.

In section 4, as for two-dimensional channels, two types of α-informations are defined and some of their properties are proved. Lastly in section 5 all these results are considered in the light of alternative definitions of conditional α-entropy and α-equivocation.

Chapter -6

In section 2 α-capacities are defined.

In section 3 the problem of maximisation of $H_\alpha(\bar{X})$ is considered for three sets of constraints on $\bar{X}$.

In section 4 as a consequence of the result obtained by maximising $H_\alpha(\bar{X})$ when the components of $\bar{X}$ have specified standard deviations, the invalidity in general of

$$H_\alpha(\bar{X}) \leq H_\alpha(X_1) + H_\alpha(X_2) + \ldots + H_\alpha(X_n) \qquad \alpha > 0$$

is shown.

In section 5.1 evaluations of α-informations, when $\bar{X}$ and $\bar{Y}$ are Gaussian, are made.

In section 5.2 α-capacities are evaluated for two cases: (i) When the channel is effected by additive Gaussian noise and each component of the input vector is distributed with mean zero and a specified standard deviation. (ii) Besides conditions in (i) input is also Gaussian.

In section 6 a generalisation of Shannon-Hartley Channel- Capacity formula is given.

In section 7 validity of various results is considered for alternative definition of conditional α-entropy and α-equivocation.

Chapter -7

In section 1 α-entropies associated with a semi-continuous multidimensional memoryless channel are defined.

In section 2 Fano's inequality for such channels is generalised.

In section 3, as a consequence of the generalisation made in section 2, a generalisation of the converse of Shannon's second fundamental theorem for semi-continuous memoryless

channels is stated and the lines, on which it can be proved for a family of extensions of such a channel, are indicated.

In section 4 these results are verified for alternative definitions of conditional α-entropy and α-equivocation.

Chapter -8
Part I

In section 1.2 α-entropies associated with different steps of a Markov chain are defined.

In section 2 some relations between these α-entropies are developed in two theorems. Theorem 1 gives two infinite sets of inequalities, each member of either set tending in the limit as $\alpha \to 1$ to the well known result

$$H_1^{(r)} = r\, H_1^{(1)}$$

where $H_1^{(r)}$ is the Shannon's entropy associated with an r step transition probability matrix of a chain. Theorem 2 establishes the validity of $H_1^{(r)} = r\, H_1^{(1)}$ when the transition probability matrix of a Markov chain is like that of a channel with symmetric noise structure.

Part II

In section 1.1 α-entropy associated with a stationary source is defined and in section 1.2 its existence is established when the underlying channel has symmetric noise structure.

In section 2 existence of other α-entropies associated with a stationary symmetric channel is established and the concepts of α-informations for such channels are developed.

3. NOTES

1. If we assume Shannon's convention that $0 \times \infty = 0$, the definitions for α-entropies given in chapters 2,3, and 4 are also valid when some of the $p(x_i)$ or conditional probabilities are zero. So this convention has been assumed in the developments of these chapters. As regards other chapters this point has separately been dealt with.

2. In all this work complete probability distributions are considered. A probability matrix $p(x_i)$ is governed by the conditions

$$0 \leq p(x_i) \leq 1, \quad \sum_i p(x_i) = 1$$

3. The meaning of $\alpha > 0$, written against a result, is that the result is true for all values of α, howsoever large , except for $\alpha = \infty$ for which its validity is not explored.

CHAPTER - 2

DISCRETE PROBABILITY SCHEMES

INTRODUCTION

Taking Renyi's entropy as the basic measure of uncertainty, various α-entropies associated with two discrete schemes are defined in section 1. Some properties like continuity, monotonicity etc. are proved in section 2. Some basic relations between α-entropies are developed in section 3 and in section 4 these are generalised for any finite number of discrete schemes. In section 5 the validity of all these results is considered in the light of another definition of conditional α-entropy and α-equivocation.

The case, when the underlying schemes are independent, has been taken into account while applying Jensen's and Hölder's inequalities.

1. α-ENTROPIES FOR TWO SCHEMES

Let there be two complete schemes

$$X = \begin{pmatrix} x_1, x_2, \ldots\ldots\ldots, x_n \\ p(x_1), p(x_2), \ldots, p(x_n) \end{pmatrix}, \quad \sum_{i=1}^{n} p(x_i) = 1$$

and

$$Y = \begin{pmatrix} y_1, y_2, \ldots\ldots\ldots, y_m \\ p(y_1), p(y_2), \ldots, p(y_m) \end{pmatrix}, \quad \sum_{j=1}^{m} p(y_j) = 1$$

Let $p(x_i, y_j)$ be the probabilities of the joint occurrence of the events x_i and y_j. The set of events $x_i y_j (1 < i < n$, $1 < j < m)$ represent joint complete scheme. Let $p(y_j/x_i)$

occurs given the occurrence of the event x_i of the scheme X so that

$$p(x_i, y_j) = p(x_i)\, p(y_j/x_i)$$

and let $p(x_i/y_j)$ be the probability of the event x_i of the scheme X occurs, given the occurrence of the event y_j of the scheme Y so that

$$p(x_i, y_j) = p(y_j)\, p(x_i/y_j)$$

These two schemes with conditional probabilities are also complete. Hence there are five α-entropies attached to the schemes X and Y :

(i) For the scheme X

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i) \qquad \alpha > 0,\ \alpha \neq 1 \tag{1}$$

(ii) For the scheme Y

$$H_\alpha(Y) = \frac{1}{1-\alpha} \log \sum_j p^\alpha(y_j) \qquad \alpha > 0,\ \alpha \neq 1 \tag{2}$$

(iii) For the joint scheme

$$H_\alpha(X,Y) = \frac{1}{1-\alpha} \log \sum_{i,j} p^\alpha(x_i, y_j) \qquad \alpha > 0,\ \alpha \neq 1 \tag{3}$$

(iv) The conditional α-entropy $H_\alpha(Y/x_i)$, calculated on the assumption that the event x_i of the scheme X has

occured is

$$H_\alpha(Y/x_i) = \frac{1}{1-\alpha} \log \sum_j p^\alpha(y_j/x_i) \quad \alpha > 0, \alpha \neq 1 \tag{4}$$

To have $H_\alpha(Y/X)$, the α-entropy attached to the conditional scheme Y/X, it is appropriate to take the expected value of (4) for all x_i ; so

$$H_\alpha(Y/X) = \frac{1}{1-\alpha} \sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i) \tag{5}$$
$$\alpha > 0, \alpha \neq 1$$

(v) Similarly $H_\alpha(X/Y)$, α-entropy attached to the conditional scheme X/Y, is given by

$$H_\alpha(X/Y) = \frac{1}{1-\alpha} \sum_j p(y_j) \log \sum_i p^\alpha(x_i/y_j) \tag{6}$$
$$\alpha > 0, \alpha \neq 1$$

It can be easily verified that in the limit, as $\alpha \to 1$, (1), (2), (3), (5) and (6) tend to the corresponding Shannon's definitions .

In the limit, as $\alpha \to \infty$, $H_\alpha(X)$ tends to

$-\log \max_i p(x_i)$ since

$$\lim_{\alpha\to\infty} H_\alpha(X) = \lim_{\alpha\to\infty} \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i)$$

$$= -\log \lim_{\alpha\to\infty} \left(\sum_i p^\alpha(x_i) \right)^{1/(1-\alpha)} \tag{7}$$

If $p(x_j)$ is the largest or one of the largest and since $0 \leqslant p(x_i) \leqslant 1$, we shall have for $\alpha > 1$

$$p^{\alpha}(x_j) \leq \sum_i p(x_i)\, p^{\alpha-1}(x_i) \leq p^{\alpha-1}(x_j)$$

or

$$\left(p^{\alpha-1}(x_j)\, p(x_j) \right)^{\frac{1}{\alpha-1}} \leq \left(\Sigma\, p^{\alpha}(x_i) \right)^{\frac{1}{\alpha-1}} \leq \left(p^{\alpha-1}(x_j) \right)^{\frac{1}{\alpha-1}}$$

or

$$p^{\frac{1}{\alpha-1}}(x_j)\, p(x_j) \leq \left(\Sigma\, p^{\alpha}(x_i)\right)^{\frac{1}{\alpha-1}} \leq p(x_j) \qquad (8)$$

So from (8) as $\alpha \to \infty$, $(\Sigma\, p^{\alpha}(x_i))^{\frac{1}{\alpha-1}}$ tends to $p(x_j)$.
Using this , (7) reduces to

$$\operatorname*{lt}_{\alpha \to \infty} H_{\alpha}(X) = -\log \max_i\ p(x_i)$$

Further since the limit of a weighted mean involving a finite number of terms is the mean of the limits of its individual terms, so we shall have

$$\operatorname*{Lt}_{\alpha \to \infty} H_{\alpha}(Y/X) = -\sum_i p(x_i) \log \max_j p(y_j/x_i)$$

Similarly

$$\operatorname*{Lt}_{\alpha \to \infty} H_{\alpha}(X,Y) = -\log \max_{i,j}\ p(x_i, y_j)$$

and

$$\operatorname*{Lt}_{\alpha \to \infty} H_{\alpha}(X/Y) = -\sum_j p(y_j) \log \max_i p(x_i/y_j)$$

2. SOME PROPERTIES

2.1 Continuity

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i)$$

$$= -\log \left(\sum_i p(x_i)\, p^{\alpha-1}(x_i)\right)^{\frac{1}{\alpha-1}}$$

$$= -\log M_{\alpha-1} \tag{9}$$

where

$$M_{\alpha-1} = \left(\sum_i p(x_i)\, p^{\alpha-1}(x_i)\right)^{\frac{1}{\alpha-1}} \tag{10}$$

Since $M_{\alpha-1}$ given by (10) is a power mean of order $\alpha-1$, $\alpha > 0$ and we know from Hardy , Littlewood and Polya (1952) that $M_{\alpha-1}$ is a continuous function of α. Hence from (9), $H_\alpha(X)$ is also a continuous function of α .

$$H_\alpha(Y/X) = \frac{1}{1-\alpha} \sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i)$$

$$= -\sum_i p(x_i) \log\left(\sum_j p(y_j/x_i)\, p^{\alpha-1}(y_j/x_i)\right)^{\frac{1}{\alpha-1}} \tag{11}$$

So $H_\alpha(Y/X)$ is a weighted mean of a finite number of continuous functions of α, hence it, itself, is a continuous function of α .

Similarly other α-entropies associated with two finite schemes are continuous functions of α.

2.2 Monotonicity

By theorem 16, Hardy, Littlewood and Polya (1952),

$M_{\alpha-1}$ given by (10) is a monotonically increasing function of α. Hence by (9), $H_\alpha(X)$ is a monotonically decreasing function of α.

It may be noted that monotonically decreasing character of $H_\alpha(X)$ can also be proved by examining the behaviour of the derivative of $H_\alpha(X)$ w.r.t. α. This procedure was adopted by Varma and Kapur (referred to in Chapter 1) for examining the behaviour of their generalisations of Renyi α -entropy.

Since by (11) $H_\alpha(Y/X)$ is a weighted mean of a finite number of monotonically decreasing functions of α, so it, itself, is a monotonically decreasing function of α.

Similarly other α-entropy associated with two finite schemes are also monotonically decreasing functions of α.

2.3 Maximality

$H_\alpha(X)$ is maximum when all the x_i are equally likely. Kapur (1968a) proved it by the technique of dynamic programming. It can also be proved by the technique of Lagrange's multipliers.

Since maximum of a weighted mean is equal to the weighted mean of the maximum values of its individual terms and each term in $H_\alpha(Y/X)$ is itself an α-entropy for fixed x so by the maximality of $H_\alpha(X)$ each of these terms is maximum when the conditional probabilities constituting it are equal.

Similarly the maximality of other α-entropies can be discussed.

2.4 An Easier Method For Successive Derivatives

In $H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i)$, let $\log \sum_i p^\alpha(x_i)$

be denoted by $f(\alpha)$, so

$$f'(\alpha) = \frac{\sum_i p^\alpha(x_i)\log p(x_i)}{\sum_i p^\alpha(x_i)} \tag{12}$$

Denoting $\frac{\sum_i p^\alpha(x_i)(\log p(x_i))^r}{\sum_i p^\alpha(x_i)}$, the weighted mean of

$(\log p(x_i))^r$ with weights $p^\alpha(x_i)$, by M_r we have

$$\frac{d M_r}{d\alpha} = \frac{\sum_i p^\alpha(x_i)(\log p(x_i))^{r+1}}{\sum_i p^\alpha(x_i)} - \frac{(\sum_i p^\alpha(x_i)(\log p(x_i))^r)(\sum_i p^\alpha(x_i)\log p(x_i))}{(\sum_i p^\alpha(x_i))^2}$$

$$= M_{r+1} - M_r M_1 \qquad \ldots(13)$$

By (12) and (13) we have

$$f'(\alpha) = M_1$$

$$f''(\alpha) = \frac{dM_1}{d\alpha} = M_2 - M_1^2$$

$$f'''(\alpha) = \frac{dM_2}{d\alpha} - \frac{d}{d\alpha} M_1^2$$

$$= M_3 - M_1 M_2 - 2M_1(M_2 - M_1^2)$$

$$= M_3 - 3 M_1 M_2 + 2 M_1^3$$

$$f^{iv}(\alpha) = M_4 - 3M_2^2 + 2M_1^4 - 4M_1 f'''(\alpha)$$

and so on

$$H_\alpha(X) = \frac{f(\alpha)}{1-\alpha}$$

$$H'_\alpha(X) = \frac{f'(\alpha)}{1-\alpha} + \frac{f(\alpha)}{(1-\alpha)^2}$$

$$= \frac{1}{1-\alpha}\left[f'(\alpha) + H_\alpha(X)\right]$$

$$H''_\alpha(X) = \frac{f''(\alpha)}{1-\alpha} + \frac{f'(\alpha)}{(1-\alpha)^2} + \frac{H'_\alpha(X)}{1-\alpha} + \frac{H_\alpha(X)}{(1-\alpha)^2}$$

$$= \frac{1}{1-\alpha}\left[f''(\alpha) + 2H'_\alpha(X)\right]$$

So in general we expect, denoting the kth derivative of $H_\alpha(X)$ by $H_\alpha^{(k)}$,

$$H_\alpha^{(k)} = \frac{1}{1-\alpha}\left[f^{(k)}(\alpha) + k\,H_\alpha^{(k-1)}\right] \qquad (14)$$

Suppose (14) is valid, so differentiating (14) again w.r.t. α we have

$$H_\alpha^{(k+1)} = \frac{f^{(k+1)}(\alpha)}{1-\alpha} + \frac{f^{(k)}(\alpha)}{(1-\alpha)^2} + \frac{K\,H_\alpha^{(k)}}{1-\alpha} + \frac{k\,H_\alpha^{(k-1)}}{(1-\alpha)^2}$$

$$= \frac{f^{k+1}(\alpha)}{1-\alpha} + \frac{k\,H_\alpha^{(k)}}{1-\alpha} + \frac{1}{1-\alpha}\left[\frac{1}{1-\alpha}\left(f_\alpha^{(k)} + k\,H_\alpha^{(k-1)}\right)\right]$$

$$H_\alpha^{(k+1)} = \frac{1}{1-\alpha}\left[f^{(k+1)}(\alpha) + (k+1)\, H_\alpha^{(k)}\right]$$

Hence in general

$$H_\alpha^{(r)} = \frac{1}{1-\alpha}\left[f^{(r)} + r\, H_\alpha^{(r-1)}\right] \tag{15}$$

(15) can be written as

$$(1-\alpha)^r H_\alpha^{(r)} = (1-\alpha)^{r-1} f^{(r)} + (1-\alpha)^{r-1}\, r\, H_\alpha^{(r-1)} \tag{16}$$

Integrating (16) from α_1 to α_2 such that $0 < \alpha_1 < \alpha_2 < \infty$, we shall have after a little simplification

$$H_{\alpha_2}^{(r-1)}(1-\alpha_2)^r = H_{\alpha_1}^{(r-1)}(1-\alpha_1)^r + \int_{\alpha_1}^{\alpha_2} (1-\alpha)^{r-1} f^{(r)}\, d\alpha \tag{17}$$

In particular putting $\alpha_1 = 1$ and $\alpha_2 = \alpha$ in (17) such that $1 < \alpha < \infty$ we have

$$H_\alpha^{(r-1)}(1-\alpha)^r = \int_1^\alpha (1-x)^{r-1} f^{(r)}(x)\, dx \tag{18}$$

Since $f^{(r)}(\alpha)$ can be easily computed with the help of (13), so consequently $H_\alpha^{(r)}$ can also be easily obtained from either the recurrence relation (15) or the integral relations (17) and (18). In particular when $\alpha = 1$ we have

a simple relation

$$H_1^{(r-1)} = -\frac{1}{r}\left[f^{(r)}(\alpha)\right] \quad \text{at } \alpha = 1$$

__Proof:__ From (15)

$$H_\alpha' = \frac{1}{1-\alpha} f' + H_\alpha$$

$$\underset{\alpha \to 1}{\mathrm{Lt}} H_\alpha' = \underset{\alpha \to 1}{\mathrm{Lt}} \frac{f' + H_\alpha}{1-\alpha} = \frac{f''(1) + H_1'}{(-1)}$$

So

$$H_1' = -\frac{1}{2} f''(1)$$

Again putting $r = 2$ in (15), taking the limit as $\alpha \to 1$ and making use of the value of H_1' obtained above we shall have

$$H_1'' = -\frac{1}{3} f'''(1)$$

So we expect

$$H_1^{(k)} = -\frac{1}{k+1} f^{(k+1)}(1) \tag{19}$$

Now putting $r = k+1$ in (15) and taking the limit as $\alpha \to 1$

$$\underset{\alpha \to 1}{\mathrm{Lt}} H_\alpha^{(k+1)} = \underset{\alpha \to 1}{\mathrm{Lt}} \frac{f^{(k+1)}(\alpha) + H_\alpha^{(k)}}{(1-\alpha)}$$

$$= \frac{f^{(k+2)}(1) + H_1^{k+1}}{(-1)} \qquad \text{(by 19)}$$

$$= -\frac{1}{k+2} f^{k+2}(1)$$

Q.E.D.

Since both $H_\alpha(Y/X)$ and $H_\alpha(X/Y)$ are weighted means of α-entropies of the sort $H_\alpha(X)$ so, similarly as above, relations involving successive derivatives of these respective conditional α-entropies can be developed.

2.5 <u>Bounds For Second Derivative</u>

$$H'_\alpha(X) = \frac{d}{d\alpha}\left(\frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i) \right)$$

$$= \frac{1}{(1-\alpha)^2} \log \sum_i p^\alpha(x_i) + \frac{1}{1-\alpha} \frac{\sum_i p^\alpha(x_i) \log p(x_i)}{\sum_i p^\alpha(x_i)}$$

$$(1-\alpha)^2 H_\alpha'(X) = \log \sum_i p^\alpha(x_i) + (1-\alpha) \frac{\sum_i p^\alpha(x_i) \log p(x_i)}{\sum_i p^\alpha(x_i)}$$

$$\frac{d}{d\alpha}\left[(1-\alpha)^2 H'_\alpha\right] = \frac{\sum_i p^\alpha(x_i) \log p(x_i)}{\sum_i p^\alpha(x_i)} - \frac{\sum_i p^\alpha(x_i) \log p(x_i)}{\sum_i p^\alpha(x_i)}$$

$$+ (1-\alpha)\left[\frac{\sum_i p^\alpha(x_i)(\log p(x_i))^2}{\sum_i p^\alpha(x_i)} - \frac{(\sum_i p^\alpha(x_i) \log p(x_i))^2}{(\sum_i p^\alpha(x_i))^2} \right]$$

$$\left[-2(1-\alpha)H'_\alpha+(1-\alpha)^2H''_\alpha\right]\left[\sum_i p^\alpha(x_i)\right]^2 = (1-\alpha)\left[\sum_i p^\alpha(x_i)\sum_i p^\alpha(x_i)(\log p(x_i))^2 - \left(\sum_i p^\alpha(x_i)\log p(x_i)\right)^2\right] \quad .. (20)$$

But by Cauchy's inequality

$$\sum_i p^\alpha(x_i) \sum_i p^\alpha(x_i)(\log p(x_i))^2 \geqslant \left(\sum_i p^\alpha(x_i)\log p(x_i)\right)^2$$

So when $0 < \alpha < 1$, by (20) we have

$$-2(1-\alpha)\, H'_\alpha + (1-\alpha)^2\, H''_\alpha \geqslant 0$$

Or

$$H''_\alpha \geqslant \frac{2\, H'_\alpha}{1-\alpha}$$

and when $\alpha > 0$ we have from (20)

$$-\,2(1-\alpha)\, H'_\alpha + (1-\alpha)^2\, H''_\alpha \leqslant 0 \quad (21)$$

or

$$H''_\alpha \leqslant \frac{2\, H'_\alpha}{1-\alpha} \quad (22)$$

When all the $p(x_i)$ are equal we have $H'_\alpha = 0$.

(20) gives a negative lower bound of H''_α for $0 < \alpha < 1$ and (21) a positive upper bound of H''_α for $\alpha > 1$. The fact that H'_α is zero when all the $p(x_i)$ are equal shows that these bounds are precise in the sense that they are actually taken on in the special case when all the $p(x_i)$

Since the derivative w.r.t. α of a weighed mean of a finite number of functions of α is the mean of their derivatives, so proceeding in the same manner, similar results for other α-entropies can be obtained.

It may be noted that the developments of the results of sections 2.4 and 2.5 are more or less the manipulations of some of the results of Beckenbach (1942) and Shniad (1948).

3. RELATIONS BETWEEN α-ENTROPIES

3.1 Between $H_\alpha(X,Y)$ and $H_\alpha(Y/X)$

(i) When $0 < \alpha < 1$

By Jensen's inequality

$$\sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i) \leq \log \sum_{i,j} p(x_i)\, p^\alpha(y_j/x_i) \qquad (23)$$

Since $p(x_i) \leq p^\alpha(x_i)$ for $0 < \alpha < 1$, so (23) can be written as

$$\sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i) \leq \log \sum_{i,j} p^\alpha(x_i, y_j) \qquad (24)$$

Multiplying (24) by $\frac{1}{1-\alpha}$, a +ve number, we have

$$H_\alpha(Y/X) \leq H_\alpha(X, Y) \qquad 0 < \alpha < 1 \qquad (25)$$

(ii) When $\alpha > 1$

Since for $\alpha > 1$, $\sum_j p^\alpha(y_j/x_i) < 1$ so we shall have

$$\sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i) \geq \sum_i p(x_i) \left\{\sum_j p^\alpha(y_j/x_i)\right\} \log\left\{\sum_j p^\alpha(y_j/x_i)\right\}$$

$$\ldots (26)$$

But $x \log x$ is a convex function of x ; so taking

$x_i = \sum_j p^\alpha(y_j/x_i)$ in (26) and using Jensen's inequality, (26) reduces to

$$\sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i) \geqslant \left[\sum_i p(x_i) \sum_j p^\alpha(y_j/x_i)\right] \log \left[\sum_i p(x_i) \sum_j p^\alpha(y_j/x_i)\right]$$

$$\geqslant \sum_{i,j} p^\alpha(x_i, y_j) \log \sum_{i,j} p^\alpha(x_i, y_j) \tag{27}$$

$$\left[\text{since } p(x_i) \geqslant p^\alpha(x_i)\right]$$

But for $\alpha > 1$, $\sum_{i,j} p^\alpha(x_i, y_j) \leqslant 1$. Let $\sum_{i,j} p^\alpha(x_i, y_j) = \beta$ (say) $\leqslant 1$,

so

$$\sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i) \geqslant \beta \log \sum_{i,j} p^\alpha(x_i, y_j) \tag{28}$$

Multiplying (28) by $\dfrac{1}{1-\alpha}$, a negative number, and using the fact that $\beta \leqslant 1$ we shall have

$$H_\alpha(Y/X) \leqslant H_\alpha(X,Y) \qquad \alpha > 1 \tag{29}$$

Combining (25) and (29) we shall have

$$H_\alpha(Y/X) \leqslant H_\alpha(X,Y) \qquad \alpha > 0,\ \alpha \neq 1 \tag{30}$$

In the limit, as $\alpha \to 1$, (30) tends to $H_1(Y/X) \leqslant H_1(X,Y)$ as already known to be true, so including this symbolically with (30) we have

$$H_\alpha(Y/X) \leqslant H_\alpha(X,Y) \qquad \alpha > 0 \tag{31}$$

Similarly we can prove

$$H_\alpha(X/Y) \leqslant H_\alpha(X,Y) \qquad \alpha > 0 \tag{32}$$

3.2 Between $H_\alpha(X,Y)$ And $H_\alpha(X)$

(i) When $0 < \alpha < 1$

$$\sum_j p^\alpha(x_i,y_j) \geqslant (\sum_j p(x_i,y_j))^\alpha = p^\alpha(x_i)(\sum_j p(y_j/x_i))^\alpha$$
$$= p^\alpha(x_i) \tag{33}$$

Summing (33) over i, taking logs and then multiplying by $\frac{1}{1-\alpha}$, a +ve no., we shall have

$$H_\alpha(X,Y) \geqslant H_\alpha(X) \qquad 0 < \alpha < 1 \tag{34}$$

(ii) When $\alpha > 1$

$$\sum_j p^\alpha(x_i,y_j) \leqslant (\sum_j p(x_i,y_j))^\alpha = p^\alpha(x_i)(\sum_j p(y_j/x_i))^\alpha = p^\alpha(x_i) \tag{35}$$

Summing (35) over i, taking logs and then multiplying by $\frac{1}{1-\alpha}$, a negative number, we have

$$H_\alpha(X,Y) \geqslant H_\alpha(X) \qquad \alpha > 1 \tag{36}$$

Combining (35) and (36) we have

$$H_\alpha(X,Y) \geqslant H_\alpha(X) \qquad \alpha > 0,\ \alpha \neq 1 \tag{37}$$

In the limit, as $\alpha \to 1$, (37) tends to $H_1(X,Y) \geqslant H_1(X)$ as known to be true, so combining this symbolically with (37) we have

$$H_\alpha(X,Y) \geqslant H_\alpha(X) \qquad \alpha > 0$$

Similarly we can prove

$$H_\alpha(X,Y) \geqslant H_\alpha(Y) \qquad \alpha > 0$$

3.3 Between $H_\alpha(Y/X)$ And $H_\alpha(Y)$

(i) When $0 < \alpha < 1$

By Jensen's inequality we have

$$\sum_i p(x_i)\log \sum_j p^\alpha(y_j/x_i) \leqslant \log \sum_i p(x_i) \sum_j p^\alpha(y_j/x_i)$$

$$= \log \sum_j \left[\sum_i p^{1-\alpha}(x_i)\, p^\alpha(x_i,y_j) \right] \qquad (38)$$

In $\sum_i p^{1-\alpha}(x_i)\; p^\alpha(x_i,y_j)$ let $1/k = \alpha$ and $1/k' = 1-\alpha$

so that $(1/k) + (1/k') = 1$ and apply Hölder's inequality for $k > 1$, th. 13, page 14, Hardy ,Littlewood and Polya (1952), we have

$$\sum_i p^{1-\alpha}(x_i)\, p^\alpha(x_i,y_j) \leqslant \left(\sum_i p(x_i)\right)^{1-\alpha} \left(\sum_i p(x_i,y_j)\right)^\alpha = p^\alpha(y_j) \quad (39)$$

Substituting (39) in (38) and dividing the result by $(1/1-\alpha)$, a positive number, we shall have

$$H_\alpha(Y/X) \leqslant H_\alpha(Y) \qquad 0 < \alpha < 1 \qquad (40)$$

(ii) When $\alpha > 1$

since $\sum_j p^\alpha(y_j/x_i) \leqslant 1$, so

$$\sum_i p(x_i)\log\sum_j p^\alpha(y_j/x_i) \geqslant \sum_i p(x_i)\left\{\sum_j p^\alpha(y_j/x_i)\right\} \log \left\{\sum_j p^\alpha(y_j/x_i)\right\} \quad (41)$$

Since $x \log x$ is a convex function of x, so taking $x_i = \sum_j p^{\alpha}(y_j/x_i)$ and applying Jensen's inequality, (41) reduces to

$$\sum_i p(x_i)\log\sum_j p^{\alpha}(y_j/x_i) \geqslant \left[\sum_i p(x_i)\sum_j p^{\alpha}(y_j/x_i)\right] \log\left[\sum_i p(x_i)\sum_j p^{\alpha}(y_j/x_i)\right]$$

$$\geqslant \sum_{i,j} p^{\alpha}(x_i)p^{\alpha}(y_j/x_i)\log\sum_{i,j} p^{1-\alpha}(x_i)p^{\alpha}(x_i,y_j) \quad \ldots. (42).$$

$$\left[\text{since } p(x_i) \geqslant p^{\alpha}(x_i) \text{ for } \alpha > 1\right]$$

But $\sum_{i,j} p^{\alpha}(x_i)p^{\alpha}(y_j/x_i) = \sum_{i,j} p^{\alpha}(x_i,y_j) \leqslant 1$, so letting

$\sum_{i,j} p^{\alpha}(x_i,y_j) = \beta \leqslant 1$, (42) reduces to

$$\sum_i p(x_i)\log \sum_j p^{\alpha}(y_j/x_i) \geqslant \beta \log \sum_j \left[\sum_i p^{1-\alpha}(x_i)\, p^{\alpha}(x_i,y_j)\right] \quad (43)$$

In $\sum_i p^{\alpha}(x_i)\, p^{\alpha}(x_i,y_j)$, let $\frac{1}{k} = \alpha$ and $\frac{1}{k'} = 1-\alpha$, so that $\frac{1}{k} + \frac{1}{k'} = 1$ and applying Hölder's inequality for $k < 1$, th. 13, page 24, Hardy, Littlewood and Polya (1952) we have

$$\sum_i p^{1-\alpha}(x_i)\, p^{\alpha}(x_i,y_j) \geqslant \Big(\sum_i p(x_i)\Big)^{1-\alpha}\Big(\sum_i p(x_i,y_j)\Big)^{\alpha} = p^{\alpha}(y_j)$$

Using this in (43) and multiplying the result so obtained by $(1/1-\alpha)$, a negative number, we shall have

$$H_{\alpha}(Y/X) \leqslant \beta H_{\alpha}(Y) \quad (44)$$

Since $\beta \leqslant 1$, so from (44) we have

$$H_\alpha(Y/X) \leqslant H_\alpha(Y) \qquad \alpha > 1 \tag{45}$$

Combining (40) and (45) we have

$$H_\alpha(Y/X) \leqslant H_\alpha(Y) \qquad \alpha > 0,\ \alpha \neq 1 \tag{46}$$

In the limit, as $\alpha \to 1$, (46) tends to $H_1(Y/X) \leqslant H_1(Y)$

as already known as Shannon's inequality, so combining this symbolically with (39) we have

$$H_\alpha(Y/X) \leqslant H_\alpha(Y) \qquad \alpha > 0 \tag{47}$$

Similarly we can prove

$$H_\alpha(X/Y) \leqslant H_\alpha(X) \qquad \alpha > 0 \tag{48}$$

3.4 <u>A Generalisation Of Khinchin's Generalisation Of Shannon's Inequality $H_1(Y/X) \leqslant H_1(Y)$</u>

Khinchin (1957) gave a generalisation of Shannon's result

$$-\sum_{i,j} p(x_i,y_j) \log p(y_j/x_i) \leqslant -\sum_j p(y_j) \log p(y_j) \tag{49}$$

that states that (49) is true if in it we sum over only a certain values of the subscript but of course over the same values of j on both the sides. We shall prove the same result for our inequality given by (46) i.e. we shall prove that the inequality in (46) is valid whether the events y

in it form a complete scheme or not. In brief it will be proved that

$$\frac{1}{1-\alpha} \sum_i p(x_i) \log \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \leq \frac{1}{1-\alpha} \log \sum_{j=1}^{*} p^{\alpha}(y_j) \qquad (50)$$

$$\alpha > 0 , \alpha \neq 1$$

where $\sum\limits_{j=1}^{*}$ denotes summation over certain values of j.

(i) <u>When $0 < \alpha < 1$</u>

By Jensen's inequality we have

$$\sum_i p(x_i) \log \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \leq \log \sum_i p(x_i) \sum_{j=1}^{*} p^{\alpha}(y_j/x_i)$$

$$= \log \sum_i \sum_{j=1}^{*} p^{1-\alpha}(x_i)\, p^{\alpha}(x_i, y_j)$$

$$= \log \sum_{j=1}^{*} p^{\alpha}(y_j) \left[\sum_i p^{\alpha}(x_i/y_j) p^{1-\alpha}(x_i) \right] \qquad (51)$$

In $\sum\limits_i p^{1-\alpha}(x_i)\, p^{\alpha}(x_i/y_j)$, let $\frac{1}{k} = \alpha$ and $\frac{1}{k'} = 1-\alpha$ so that $\frac{1}{k} + \frac{1}{k'} = 1$ and apply Hölder's inequality for $k > 1$, we shall have

$$\sum_i p^{1-\alpha}(x_i)\, p^{\alpha}(x_i/y_j) \leq \Big(\sum_i p(x_i) \Big)^{1-\alpha} \Big(\sum_i p(x_i/y_j) \Big)^{\alpha} = 1$$

Let $\sum_i p^{1-\alpha}(x_i)\, p^{\alpha}(x_i/y_j) = \beta_j \leq 1$ and $\beta = \max_j \beta_j$.

Substituting this in (51) we shall have

$$\sum_i p(x_i) \log \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \leq \log \sum_{j=1}^{*} p^{\alpha}(y_j) + \log \beta \qquad (52)$$

Multiplying (52) by $(1/1-\alpha)$, a +ve number, and using the fact that $\beta \leq 1$ we shall have (50) for $0 < \alpha < 1$.

<u>When $\alpha > 1$</u>

Since $\sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \leq 1$, so

$$\sum_i p(x_i) \log \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \geqslant \sum_i p(x_i) \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \log \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \qquad (53)$$

Since $x \log x$ is a convex function of x, so taking $x_i = \sum_{j=1}^{*} p^{\alpha}(y_j/x_i)$ and applying Jensen's inequality, (53) reduces to

$$\sum_i p(x_i) \log \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \geqslant \left[\sum_i p(x_i) \sum_{j=1}^{*} p^{\alpha}(y_j/x_i)\right] \log\left[\sum_i p(x_i) \sum_{j=1}^{*} p^{\alpha}(y_j/x_i)\right] \qquad \ldots (54)$$

Since $\sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \leq 1$, letting $\min \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) = \beta \leq 1$, so this and a little manipulation would reduce (54) to

$$\sum_i p(x_i) \log \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \geqslant \beta \log \sum_{j=1}^{*} p^{\alpha}(y_j) \left\{\left[\sum_i p^{\alpha}(x_i/y_j) p^{1-\alpha}(x_i)\right]\right. \qquad (55)$$

In $\sum_i p^{1-\alpha}(x_i)\, p^{\alpha}(x_i/y_j)$, let $\frac{1}{k} = \alpha$, $\frac{1}{k'} = 1 - \alpha$ and apply Hölder's inequality for $k < 1$, we shall have

$$\sum_i p^{1-\alpha}(x_i)\, p^{\alpha}(x_i/y_j) \geqslant \Big(\sum_i p(x_i)\Big)^{1-\alpha} \Big(\sum_i p(x_i/y_j)\Big)^{\alpha} = 1$$

Let $\sum_i p^{1-\alpha}(x_i)\, p^{\alpha}(x_i/y_j) = \gamma_j \geqslant 1$ and $\gamma = \min \gamma_j$. so substituting this in (55) we shall have

$$\sum_i p(x_i) \log \sum_{j=1}^{*} p^{\alpha}(y_j/x_i) \geqslant \beta \log\left[\sum_{j=1}^{*} p^{\alpha}(y_j) + \log \gamma\right] \qquad (56)$$

Multiplying (56) by $(1/1-\alpha)$, a -ve number, and using the fact that $\beta \leqslant 1$ and $\gamma \geqslant 1$ we shall have (50) for $\alpha > 1$.

In the limit , as $\alpha \to 1$, (50) tends to Khinchin's generalisation.

Similarly proceeding as above we can prove

$$\frac{1}{1-\alpha} \sum_{j=1}^{m} p(y_j) \log \sum_{i=1}^{*} p^{\alpha}(x_i/y_j) \leqslant \frac{1}{1-\alpha} \log \sum_{i=1}^{*} p^{\alpha}(x_i) \qquad (57)$$

$\alpha > 0,\ \alpha \neq 1$

In the limit, as $\alpha \to 1$, (57) tends to Khinchin's generalisation of $H_1(X/Y) \leqslant H_1(X)$

3.5 Study of Relations Between $H_\alpha(X,Y)$ And $H_\alpha(X) + H_\alpha(Y/X)$

Two infinite sets of inequalities, one for the range $0 < \alpha < 1$ and one for $\alpha > 1$, each tending in the limit as $\alpha \to 1$ to Shannon's result $H_1(X,Y) = H_1(X) + H_1(Y/X)$ have been developed. Two bounds for $H_\alpha(X,Y)$, one upper for $0 < \alpha < 1$ and are lower for $\alpha > 1$, have also been developed.

Case (i) When $0 < \alpha < 1$

$$\log \sum_{i,j} p^{\alpha}(x_i,y_j) = \log \sum_{i,j} p^{\alpha-\frac{1}{k}}(x_i)\, p^{1/k}(x_i)\, p^{\alpha}(y_j/x_i)$$

Where k is so chosen that $0 < k < 1$

$$\log \sum_{i,j} p^{\alpha}(x_i,y_j) = \log \sum_{i} \left[\left\{ p^{\alpha-1/k}(x_i) \right\} \left\{ \sum_j p^{1/k}(x_i) p^{\alpha}(y_j/x_i) \right\} \right] \quad (58)$$

Choosing k' subject to $\frac{1}{k} + \frac{1}{k'} = 1$ and taking $\sum_j p^{1/k}(x_i)\, p^{\alpha}(y_j/x_i) = a_i$ and $p^{\alpha-1/k}(x_i) = b_i$ in $\sum_i a_i b_i \geqslant \left(\sum_i a_i^k\right)^{1/k} \left(\sum_i b_i^{k'}\right)^{1/k'}$, the result of Hölder inequality for $0 < k < 1$, (58) reduces to

$$\log \sum_{i,j} p^{\alpha}(x_i,y_j) \geqslant \log \left[\sum_i \left\{ p^{\alpha-1/k}(x_i) \right\}^{k'} \right]^{1/k'} + \log \left[\sum_i \left\{ \sum_j p^{1/k}(x_i) p^{\alpha}(y_j/x_i) \right\}^{k} \right]^{1/k}$$

$$= \frac{1}{k'} \log \sum_i p^{(\alpha-1/k)k'}(x_i) + \frac{1}{k} \log \sum_i p(x_i) \left[\sum_j p^{\alpha}(y_j/x_i) \right]^{k} \quad (59)$$

But by Jensen's inequality

$$\log \sum_i p(x_i) \left[\sum_j p^{\alpha}(y_j/x_i) \right]^{k} \geqslant k \sum_i p(x_i) \log \sum_j p^{\alpha}(y_j/x_i) \quad (60)$$

Using (60), (59) reduces to

$$\log \sum_{i,j} p^{\alpha}(x_i, y_j) \geqslant \frac{1}{k'} \log \sum_{i} p^{(\alpha - \frac{1}{k})k'}(x_i) + \sum_{i} p(x_i) \log \sum_{j} p^{\alpha}(y_j/x_i) \quad (61)$$

Multiplying (61) by $(1/1-\alpha)$, a positive number, and making use of the fact that $(1-\alpha)k' = (1/k + 1/k' - \alpha)k' = 1 - (\alpha - 1/k)k'$, we shall have

$$H_{\alpha}(X, Y) \geqslant H_{(\alpha - 1/k)k'}(X) + H_{\alpha}(Y/X) \quad (62)$$

$$0 < \alpha < 1, \; 0 < k < 1 \text{ and } 1/k + 1/k' = 1$$

Under the conditions for the validity of (62), $(\alpha - 1/k)k'$ lies between 1 and ∞ since

$$(\alpha - 1/k)k' = \frac{1 - \alpha k}{1 - k} \quad (63)$$

For $0 < \alpha < 1$, $0 < k < 1$ we have

$$\alpha k < k$$

or $$\alpha k - 1 < k - 1$$

or $$\frac{1 - \alpha k}{1 - k} > 1,$$ so from (63) $(\alpha - 1/k)k' > 1$.

So from (62) accrues an infinite number of inequalities connecting $H_{\alpha}(X,Y)$ and $H_{\alpha}(Y/X)$ for $0 < \alpha < 1$ to $H_{\alpha}(X)$ for $\alpha > 1$. As $\alpha \to 1$, all these inequalities tend to the Shannon's equality $H_1(X,Y) = H_1(X) + H_1(Y/X)$.

Case (ii) When $0 < \alpha < 1$

Rewriting (58) on the assumption that $k > 1$, we have

$$\log \sum_{i,j} p^{\alpha}(x_i,y_j) = \log\left[\sum_i \left\{p^{(\alpha-1/k)}(x_i)\right\}\left\{\sum_j p^{1/k}(x_i)p^{\alpha}(y_j/x_i)\right\}\right] \quad (64)$$

Choosing k' such that $1/k + 1/k' = 1$ and applying Hölder inequality for $k > 1$, (64) reduces to

$$\log \sum_{i,j} p^{\alpha}(x_i,y_j) \leq \log\left[\sum_i\left\{p^{(\alpha-1/k)}(x_i)\right\}^{k'}\right]^{1/k'} + \log\left[\sum_i\left\{\sum_j p^{1/k}(x_i)p^{\alpha}(y_j/x_i)\right\}^{k}\right]^{\frac{1}{k}} \quad \ldots \quad (65)$$

Multiplying (65) by $(1/1-\alpha)$, a +ve number, we have

$$H_{\alpha}(X,Y) \leq H_{(\alpha-1/k)k'}(X) + \frac{1}{(1-\alpha)k}\log \sum_i p(x_i)\left[\sum_j p^{\alpha}(y_j/x_i)\right]^{k} \quad (66)$$

$$0 < \alpha < 1, \ k > 1 \text{ and } 1/k + 1/k' = 1$$

Taking limits of (66) as $k \to \infty$ we have

$$\operatorname*{Lt}_{k \to \infty} H_{(\alpha-1/k)k'}(X) = H_{\alpha}(X)$$

$$\text{and } \operatorname*{Lt}_{k \to \infty} \frac{1}{(1-\alpha)k}\log \sum_i p(x_i)\left[\sum_j p^{\alpha}(y_j/x_i)\right]^{k} \quad (67)$$

and

$$\underset{k \to \infty}{\mathrm{Lt}} \frac{1}{(1-\alpha)k} \log \sum_i p(x_i) \left[\sum_j p^{\alpha}(y_j/x_i) \right]^k \tag{67}$$

$$= \underset{k \to \infty}{\mathrm{Lt}} \sum_i \frac{\left[p(x_i) \left\{ \sum_j p^{\alpha}(y_j/x_i) \right\}^k \right] \frac{1}{1-\alpha} \log \sum_j p^{\alpha}(y_j/x_i)}{\sum_i \left[p(x_i) \left\{ \sum_j p^{\alpha}(y_j/x_i) \right\}^k \right]} \tag{68}$$

Since the expression on the R.H.S. of (68), whose limit, as $k \to \infty$, is to be taken, is a weighted mean of the quantities $(1/1-\alpha) \log \sum_j p^{\alpha}(y_j/x_i)$, so it should be less than $\max_i \left[(1/1-\alpha) \log \sum_j p^{\alpha}(y_j/x_i) \right]$. Hence using this and (67), (66), in the limit as $k \to \infty$, reduces to

$$H_{\alpha}(X,Y) \leq H_{\alpha}(X) + \max_i \left\{ (1/1-\alpha) \log \sum_j p^{\alpha}(y_j/x_i) \right\}$$

or equivalently

$$H_{\alpha}(X,Y) \leq H_{\alpha}(X) + \max_i . H_{\alpha}(Y/x_i) \tag{69}$$

$$0 < \alpha < 1$$

Note that (69) gives an upper bound for $H_{\alpha}(X,Y)$ for $0 < \alpha < 1$.

Case (i) When $\alpha > 1$,

Note that in deriving (61) no condition on α was made, so multiplying it by $(1/1-\alpha)$, a negative number, we have

$$H_{\alpha}(X,Y) \leq H_{(\alpha-1/k)k'}(X) + H_{\alpha}(Y/X) \tag{70}$$

$$\alpha > 1,\ 0 < k < 1 \text{ and } 1/k + 1/k' = 1$$

We observe that $(\alpha - 1/k)\,k' < 1$. Under the given conditions for which (70) is valid, $(\alpha - 1/k)k' = \frac{\alpha k-1}{k-1}$ and $\alpha k > k$. Now there are two cases (i) $\alpha k > 1$, (ii) $\alpha k < 1$. In case (i) $(\alpha - 1/k)k'$ is negative and in case (ii) it lies between 0 and 1.

So from (70) accrues an infinite number of inequalities connecting $H_\alpha(X,Y)$ and $H_\alpha(Y/X)$ for $\alpha > 1$, $\alpha \neq 1$ to $H_\alpha(X)$ for $0 < \alpha < 1$. As $\alpha \to 1$, all these inequalities tend to the Shannon's result $H_1(X,Y) = H_1(X) + H_1(Y/X)$.

Case (ii) When $\alpha > 1$

Since in deriving (65), no condition on α was made, so multiplying (65) by $(1/1-\alpha)$, a negative number, we have

$$H_\alpha(X,Y) \geqslant H_{(\alpha-1/k)k'}(X) + \frac{1}{(1-\alpha)k}\log \sum_i p(x_i)\left[\sum_j p^\alpha(y_j/x_i)\right]^k \tag{71}$$

Taking limits of (71) as $k \to \infty$ we have

$$\underset{k\to\infty}{\mathrm{Lt}}\ H_{(\alpha-1/k)k'}(X) = H_\alpha(X) \tag{72}$$

and

$$\underset{k\to\infty}{\mathrm{Lt}}\ \frac{1}{(1-\alpha)k}\log \sum_i p(x_i)\left[\sum_j p^\alpha(y_j/x_i)\right]^k$$

$$= \underset{k\to\infty}{\mathrm{Lt}}\ \sum_i \frac{\left[p(x_i)\left\{\sum_j p^\alpha(y_j/x_i)\right\}^k\right]\frac{1}{1-\alpha}\log\sum_j p^\alpha(y_j/x_i)}{\sum_i\left[p(x_i)\left\{\sum_j p^\alpha(y_j/x_i)\right\}^k\right]}$$

$$\geqslant \min_i\left\{\frac{1}{1-\alpha}\log\sum_j p^\alpha(y_j/x_i)\right\} = \min_i\ H_\alpha(Y/x_i) \tag{73}$$

Using (72) and (73), (71) reduces to

$$H_\alpha(X,Y) \geqslant H_\alpha(X) + \min_i H_\alpha(Y/x_i) \qquad \alpha > 1 \qquad (74)$$

(74) gives us a lower bound for $H_\alpha(X,Y)$ for $\alpha > 1$.
Similarly, proceeding as above, we shall have for $0 < \alpha < 1$

$$H_\alpha(X,Y) \geqslant H_{(\alpha-1/k)k'}(Y) + H_\alpha(X/Y)$$

$$0 < k < 1, \quad 1/k + 1/k' = 1$$

$$H_\alpha(X,Y) \leqslant H_\alpha(Y) + \max_j H_\alpha(X/y_j)$$

and for $\alpha > 1$

$$H_\alpha(X,Y) \leqslant H_{(\alpha-1/k)k'}(Y) + H_\alpha(X/Y)$$

$$0 < k < 1, \quad 1/k + 1/k' = 1$$

$$H_\alpha(X,Y) \geqslant H_\alpha(Y) + \min_j H_\alpha(X/y_j)$$

3.6 <u>A Study of The Validity Of "$H_\alpha(X,Y) = H_\alpha(X) + H_\alpha(Y/X)$"</u>

It will be proved that when the channel is either symmetric or binary erasure, the result $H_\alpha(X,Y) = H_\alpha(X)+H_\alpha(Y/X)$ for $\alpha > 0$ will hold but it is not true in general.

<u>When the channel is symmetric</u>

Since the channel is symmetric if each of its rows and each of its columns consists of same set of elements, may not be in the same order, so for such a channel $\sum_j p^\alpha(y_j/x_i)$ = a constant,

$i = 1, 2, \ldots, n$. Hence

$$H_\alpha(X,Y) = \frac{1}{1-\alpha} \log \sum_{i,j} p^\alpha(x_i, y_j)$$

$$= \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i) \sum_j p^\alpha(y_j/x_i)$$

$$= \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i) + \frac{1}{1-\alpha} \log \sum_j p^\alpha(y_j/x_i)$$

$$= \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i) + \frac{1}{1-\alpha} \sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i)$$

$$= H_\alpha(X) + H_\alpha(Y/X) \qquad (75)$$

<u>When the channel is binary erasure</u>

Again for such a channel $\sum_j p^\alpha(y_j/x_i)$ = a constant for $i = 1, 2, \ldots, n$ and the truth of $H_\alpha(X,Y) = H_\alpha(X) + H_\alpha(Y/X)$ can similarly be checked.

<u>Note</u> that validity of $H_\alpha(X,Y) = H_\alpha(X) + H_\alpha(Y/X)$ does not imply the validity of $H_\alpha(X,Y) = H_\alpha(Y) + H_\alpha(X/Y)$

<u>An Example</u>

In general

$$H_\alpha(X,Y) \neq H_\alpha(X) + H_\alpha(Y/X) \neq H_\alpha(Y) + H_\alpha(X/Y) \quad \alpha > 1,\ \alpha \neq 1 \qquad (76)$$

Suppose that the symbols 0 and 1 are transmitted with respective probabilities 1/3 and 2/3 and received as either ~~as~~ 0 or 1 or 2 with the following noise characteristic of the channel:

X \ Y	0	1	2
0	1/4	1/2	1/4
1	0	2/5	3/5

Calculations would give us

$p(Y = 0) = 1/12$, $p(Y=1) = 13/30$ and $p(Y=2) = 29/60$

$$P\{X, Y\} = \begin{bmatrix} 1/12 & 1/6 & 1/12 \\ 0 & 4/15 & 2/5 \end{bmatrix}$$

and calculations made by taking logs to the base 10 would give

$H_2(X,Y)$	$= .5642$	$H_{1/2}(X,Y)$	$= .6382$
$H_2(X)$	$= .2552$	$H_{1/2}(X)$	$= .2884$
$H_2(Y)$	$= .3682$	$H_{1/2}(Y)$	$= .4310$
$H_2(Y/X)$	$= .3313$	$H_{1/2}(Y/X)$	$= .3525$
$H_2(X/Y)$	$= .2521$	$H_{1/2}(X/Y)$	$= .2460$

(76) can be easily verified for $\alpha = 2$ and $\alpha = 1/2$.

4. FOR FINITE NO. OF DISCRETE SCHEMES

Consider r finite discrete schemes specified respectively by the discrete random variables $X_1, \ldots, X_r$. Let the product scheme obtained by joining these r schemes be denoted by $X_1 \otimes \ldots \otimes X_r$. In this product scheme $X_1 \otimes \ldots \otimes X_r$, following are a few generalisations of the results developed in section 3 :

(i) the inequality in (47) will get generalised to inequalities of the form

$$H_\alpha(X_r/X_1, X_2, \ldots, X_{r-1}) \leq H_\alpha(X_r/X_2, \ldots, X_{r-1}) \;, \alpha > 0 \quad (77)$$

(ii) For $0 < \alpha < 1$, the inequalities in (62) will get generalised to inequalities of the form

$$H_\alpha(X_1 \ldots, X_r) \geq H_{(\alpha-1/k_1)k_1'}(X_1) + H_{(\alpha-1/k_2)k_2'}(X_2/X_1) + \ldots$$

$$+ H_{(\alpha-1/k_{r-1})k_{r-1}'}(X_{r-1} / X_1, \ldots, X_{r-2})$$

$$+ H_\alpha(X_r / X_1, X_2, \ldots X_{r-1}) \tag{78}$$

where $0 < k_1, \ldots, k_{r-1} < 1$ and $1/k_1 + 1/k_1' = 1, \ldots, 1/k_{r-1} + 1/k'_{r-1} = 1$ and the inequality in (69) to inequalities of the form

$$H_\alpha(X_1, \ldots, X_r) \leq H_\alpha(X_1) + \max_{x_1} H_\alpha(X_2/x_1)$$

$$+ \max_{x_1, x_2} H_\alpha(X_3/x_1, x_2)$$

$$+ \max_{x_1, \ldots, x_{r-1}} H_\alpha(X_r / x_1, \ldots, x_{r-1}) \tag{79}$$

(iii) For $\alpha > 1$, the inequalities in (70) will get generalised to inequalities of the form

$$H_\alpha(X_1, \ldots, X_r) \leq H_{(\alpha-1/k_1)k_1'}(X_1) + H_{(\alpha-1/k_2)k_2'}(X_2/X_1) + \ldots$$

$$+ H_{(\alpha-1/k_{r-1})k'_{r-1}}(X_{r-1}/X_1, X_2, \ldots, X_{r-2})$$

$$+ H_\alpha(X_r/X_1, X_2, \ldots, X_r-1) \tag{80}$$

where $0 < k_1, \ldots, k_{r-1} < 1$ and $1/k_1 + 1/k_1' = 1, \ldots, 1/k_{r-1} + 1/k_{r-1}' = 1$ and the inequality in (74) to inequalities of the form

$$H_\alpha(X_1,\ldots,X_r) \geqslant H_\alpha(X_1) + \min_{x_1} H_\alpha(X_2/x_1)$$

$$+ \min_{x_1,x_2} H_\alpha(X_3/x_1,x_2) + \ldots$$

$$+ \min_{x_1,\ldots,x_{r-1}} H_\alpha(X_r/x_1,\ldots,x_{r-1}) \qquad (81)$$

Proof of (77) :

On the assumption that an event $(x_2,\ldots,x_{r-1})$ of the product space $X_2 \otimes \ldots \otimes X_{r-1}$ has occurred, let the probabilities of the events of the type (x_1,x_r) belonging to $X_1 \otimes X_r$ be given by

$$p(x_1,x_r/x_2,\ldots,x_{r-1}) = q(x_1,x_r) \qquad (82)$$

and the schemes given by the random variables X_1 and X_r become the schemes given by the random variables X_1' and X_r' with probabilities of the type

$$p(x_1/x_2,\ldots,x_{r-1}) = q(x_1) \qquad (83)$$

and

$$p(x_r/x_2,\ldots,x_{r-1}) = q(x_r) \qquad (84)$$

By (46) for $\alpha > 0$, $\alpha \neq 1$, we have

$$H_\alpha(X_r'/X_1') \leq H_\alpha(X_r') \qquad (85)$$

Using (83) and (84), the L.H.S. of (85) is given by

$$H_\alpha(X_r'/X_1') = \frac{1}{1-\alpha} \sum_{X_1'} q(x_1) \log \sum_{X_r'} q^\alpha(x_r/x_1) \qquad (86)$$

Using (82) and (83) we have

$$q(x_r/x_1) = \frac{q(x_r,x_1)}{q(x_1)} = \frac{p(x_r,x_1/x_2,\ldots,x_{r-1})}{p(x_1/x_2,\ldots,x_{r-1})}$$

$$= p(x_r/x_1,x_2,\ldots,x_{r-1}) \tag{87}$$

Using (87) and (83) , (86) reduces to

$$H_\alpha(X_r'/X_1') = \frac{1}{1-\alpha}\sum_{X_1} p(x_1/x_2,\ldots,x_{r-1})\log\sum_{X_r} p^\alpha(x_r/x_1,x_2,\ldots,x_{r-1}) \quad \ldots (88)$$

Using (84)

$$H_\alpha(X_r') = \frac{1}{1-\alpha}\log\sum_{X_r} p^\alpha(x_r/x_2,\ldots,x_{r-1}) \tag{89}$$

Using (88) and (89), (86) reduces to

$$\frac{1}{1-\alpha}\sum_{X_1} p(x_1/x_2,\ldots,x_{r-1})\log\sum_{X_r} p^\alpha(x_r/x_1,x_2,\ldots,x_{r-1})$$

$$\leqslant \frac{1}{1-\alpha}\log\sum_{X_r} p^\alpha(x_r/x_2,\ldots,x_{r-1}) \tag{90}$$

Multiply (90) by $p(x_2,\ldots,x_{r-1})$ and summing over $X_2 \otimes \ldots \otimes X_{r-1}$ we shall have

$$H_\alpha(X_r/X_1,X_2,\ldots,X_{r-1})$$

$$\leqslant H_\alpha(X_r/X_2,\ldots,X_{r-1}) \tag{91}$$

$$\alpha > 0,\ \alpha \neq 1$$

In the limit, as $\alpha \to 1$, (91) tends to the known result $H_1(X_r/X_1,\ldots,X_{r-1}) \leqslant H_1(X_r/X_2,\ldots,X_{r-1})$; combining this

with (91) we shall have (77).

Q.E.D.

<u>Outline of the Proof for (78)</u> :

Applying the procedure followed to get (62), we shall have

$$H_\alpha(X_1,\ldots,X_r) \geqslant H_{(\alpha-1/k_1)k'}(X_1) + H_\alpha(x_2,\ldots,X_r/X_1) \qquad (92)$$

where $0 < k_1 < 1$ and $1/k_1 + 1/k_1' = 1$. But in (92)

$$H_\alpha(X_2,\ldots,X_r/X_1) = (1/1-\alpha)\sum_{X_1} p(x_1)\log \sum_{X_2\otimes\ldots\otimes X_r} p^\alpha(x_2,\ldots,x_r/x_1)$$

Writing

$$\log \sum_{X_2\otimes\ldots\ldots\otimes X_r} p^\alpha(x_2,\ldots,x_r/x_1)$$

$$= \log \sum_{X_2} \left\{p(x_2/x_1)\right\}^{\alpha-1/k_2} \left\{\sum_{X_3\otimes\ldots\ldots\otimes X_r} p^{1/k_2}(x_2/x_1)\, p^\alpha(x_3,\ldots,x_r/x_1,x_2)\right\}$$

where $0 < k_2 < 1$, choosing k_2' so that $1/k_2 + 1/k_2' = 1$, applying Hölder inequality and following the procedure adopted for getting (62) we shall have

$$H_\alpha(X_2,\ldots,X_r/x_1) \geqslant H_{(\alpha-1/k_2)k_2'}(X_2/x_1)$$
$$+(1/1-\alpha)\sum_{X_2} p(x_2/x_1)\log\sum_{X_3\otimes\ldots\otimes X_r} p^\alpha(x_3,\ldots,x_r/x_1,x_2) \qquad (93)$$

Multiplying (93) by $p(x_1)$ and summing over X_1 we shall have

$$H_\alpha(X_2,\ldots,X_r/X_1) \geqslant H_{(\alpha-1/k_2)k_2'}(X_2/X_1) + H_\alpha(X_3,\ldots,X_r/X_1,X_2) \qquad (94)$$

Combining (94) with (92) and applying to $H_\alpha(X_3,\ldots,X_r/X_1,X_2)$ the procedure for getting (94) for another r-3 times, we shall have (78).

<u>Outline of the Proof for(79):</u>

$$H_\alpha(X_1,\ldots,X_r) = \frac{1}{1-\alpha} \log \sum_{X_1\otimes\ldots\otimes X_r} p^\alpha(x_1,\ldots,x_r)$$

$$= \frac{1}{1-\alpha} \log \sum_{X_1\otimes\ldots\otimes X_r} p^{\alpha-1/k_1}(x_1)\, p^{1/k_1}(x_1)\, p^\alpha(x_2,\ldots,x_r/x_1) \quad (95)$$

where $k_1 > 1$. Following the same procedure as used for getting (69), from (95) we shall have

$$H_\alpha(X_1,\ldots,X_r) \leq H_\alpha(X_1) + \max_{x_1} H_\alpha(X_2,\ldots,X_r/x_1) \quad (96)$$

But in (96)

$$H_\alpha(X_2,\ldots,X_r/x_1) = \frac{1}{1-\alpha} \log \sum_{X_2\otimes\ldots\otimes X_r} p^\alpha(x_2,\ldots,x_r/x_1)$$

$$= \frac{1}{1-\alpha} \log \sum_{X_2\otimes\ldots\otimes X_r} p^{\alpha-1/k_1}(x_2/x_1)\, p^{1/k_1}(x_2/x_1)\, p^\alpha(x_3,\ldots,x_r/x_1,x_2) \quad \ldots (97)$$

Again applying to (97) the procedure adopted to get (69) from (64) we shall have

$$\max_{x_1} H_\alpha(X_2,\ldots,X_r/x_1) \leq \max_{x_1} \left[H_\alpha(X_2/x_1) + \max_{x_2} H_\alpha(X_3,\ldots,X_r/x_1,x_2) \right]$$

$$= \max_{x_1} H_\alpha(X_2/x_1) + \max_{x_1,x_2} H_\alpha(X_3,\ldots,X_r/x_1,x_2) \quad (98)$$

Combining (96) with (98) and applying to $\max_{x_1,x_2} H_\alpha(X_3,\ldots,X_r/x_1,x_2)$

have (79).

(80) and (81) can similarly be proved by repeating r-1 times the procedure adopted for getting (70) and (74).

5. ALTERNATIVE DEFINITION OF CONDITIONAL α-ENTROPY AND VERIFICATIONS.

There is another way of defining conditional α-entropy. Instead of taking the weighted mean of $H_\alpha(Y/x_i)$, the way we did for deriving (5) from (4), we can also consider again the Kolmogorov-Nagumo generalised mean of the 'information contained' variable $H_\alpha(Y/x_i)$ w.r.t. $g(x)= 2^{(1-\alpha)x}$ $\alpha>0, \alpha \neq 1$ as we implicitly did for getting $H_\alpha(Y/x_i)$ or Renyi (1961) did for getting $H_\alpha(X)$. So if we do this way, the conditional α-entropy would be

$$H_\alpha(Y/X) = g^{-1}\left[\sum_i p(x_i)\, 2^{(1-x)\, H_\alpha(Y/x_i)}\right] \tag{99}$$

But by (5)

$$H_\alpha(Y/x_i) = \frac{1}{1-\alpha} \log \sum_j p^\alpha(y_j/x_i) \, ,$$

so substituting this in (99) and simplifying we have

$$H_\alpha(Y/X) = \frac{1}{1-\alpha} \log \left[\sum_i p(x_i) \left\{\sum_j p^\alpha(y_j/x_i)\right\}\right] \tag{100}$$

$$\alpha > 0,\ \alpha \neq 1$$

Similarly following these lines α-equivocation would be

$$H_\alpha(X/Y) = \frac{1}{1-\alpha} \log \left[\sum_j p(y_j) \left\{\sum_i p^\alpha(x_i/y_j)\right\}\right] \tag{101}$$

In the limit, as $\alpha \to 1$, both (100) and (101) tend to the corresponding Shannon's definitions. But by Jensen's inequality

$$\log \left[\sum_i p(x_i) \left\{ \sum_j p^{\alpha}(y_j/x_i) \right\} \right] \geq \sum_i p(x_i) \log \sum_j p^{\alpha}(y_j/x_i),$$

so for $0 < \alpha < 1$

$$\frac{1}{1-\alpha} \sum_i p(x_i) \log \sum_j p^{\alpha}(y_j/x_i) \leq \frac{1}{1-\alpha} \log \left[\sum_i p(x_i) \left\{ \sum_j p^{\alpha}(y_j/x_i) \right\} \right] \quad \ldots \quad (102)$$

and for $\alpha > 1$

$$\frac{1}{1-\alpha} \sum_i p(x_i) \log \sum_j p^{\alpha}(y_j/x_i) \geq \frac{1}{1-\alpha} \log \left[\sum_i p(x_i) \left\{ \sum_j p^{\alpha}(y_j/x_i) \right\} \right] \quad \ldots \quad (103)$$

By (102) and (103) conditional α-entropy given by (5) is less for $0 < \alpha < 1$ and more for $\alpha > 1$, than that given by (100). Similar interpretation can be had for (6) by comparing it with (101) by Jensen's inequality.

Results which involve specifically the definitions of $H_{\alpha}(Y/X)$ or $H_{\alpha}(X/Y)$ and given by the following equations are also true if we use the alternative definitions of $H_{\alpha}(Y/X)$ and $H_{\alpha}(X/Y)$ given by (100) and (101) respectively :

All the results, except those given by (62) and (70) and their generalisations given respectively by (78) and (80), are also true for these alternative definitions of conditional α-entropy and α-equivocation. These can be proved for the definitions given by (100) and (101) by simple manipulations

of the procedures adopted to prove them for definitions given by (5) and (6). For instance the result in (47) for $H_\alpha(Y/X)$, given by (100), can be proved as follows :

For $0 < \alpha < 1$, comparing (38) and the definitions of $H_\alpha(Y/X)$ given by (5) and (100) and retracing the steps for obtaining (40) from (38), the result follows.

For $\alpha > 1$, by Jensen's inequality we have

$$\log \sum_i p(x_i) \sum_j p^\alpha(y_j/x_i) \geqslant \sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i)$$

and further retracing the steps used for getting (45) from (41) we shall have the result. Or else

$$\log \sum_{i,j} p(x_i)\, p^\alpha(y_j/x_i) = \log \sum_j \left[\sum_i p^{1-\alpha}(x_i) p^\alpha(x_i, y_j) \right] \quad (104)$$

By the argument used for getting (44) from (43) we have

$$\sum_i p^{1-\alpha}(x_i)\, p^\alpha(x_i, y_j) \geqslant p^\alpha(y_j) \quad (105)$$

Using (105), (104) reduces to

$$\log \sum_{i,j} p(x_i)\, p^\alpha(y_j/x_i) \geqslant \log \sum_j p^\alpha(y_j) \quad (106)$$

Multiplying (106) by $\dfrac{1}{1-\alpha}$, a negative number, we shall have the result.

CHAPTER - 3

DISCRETE MEMORYLESS CHANNELS

1. INTRODUCTION

Since α-entropies defined in section 1 of chapter 2 are associated with two finite discrete schemes, so these are also the α-entropies associated with a two-dimensional discrete memoryless channel with input and output specified respectively by X and Y or a communication system with two parts. Translating these for such a channel we have

(i) $H_\alpha(X)$ is defined for the input X.

(ii) $H_\alpha(Y)$ is defined for the output Y.

(iii) $H_\alpha(X,Y)$ is defined for the communication system as a whole.

(iv) $H_\alpha(Y/X)$ is defined about the output where it is known that X is transmitted and may be termed as conditional α-entropy.

(v) $H_\alpha(X/Y)$ is defined about the input where it is known that Y is received and may be termed as α-equivocation.

The results given by (47), (48) and (76) of chapter 2 make us define two types of α- informations processed in the channel and hence two types of α-capacities, α-redundancies etc. and these concepts are defined in section 2. In section 3, α-information, when, corresponding to a single input, a number of outputs are received, is defined and the available result on the additivity of mutual information is generalised. In section 4, following McGill (1954) α-information of several alphabets is defined and expressed in terms of α- entropies.

information through cascaded channels is generalised. In section 6, behaviour and α-capacities of five varieties of channels are considered. In section 7, the validity of the results of this chapter is examined in the light of second definition of conditional α-entropy developed in section 5 of chapter 2.

2. α-INFORMATIONS, α-CAPACITIES, α-REDUNDANCIES AND α-EFFICIENCIES

2.1 α-Informations

We may define α-information, $\alpha > 0$, $\alpha \neq 1$, conveyed about X by Y as

$$I_\alpha(X;\, Y) = H_\alpha(X) - H_\alpha(X/Y) \tag{1}$$

and α-information, $\alpha > 0$, $\alpha \neq 1$, conveyed about Y by X as

$$I_\alpha(Y\,;X) = H_\alpha(Y) - H_\alpha(Y/X) \tag{2}$$

Note that, by (76) of chapter 2, in general $I_\alpha(X;Y) \neq I_\alpha(Y;X)$ and, by (47) and (48) of chapter 2, both $I_\alpha(Y;X)$ and $I_\alpha(X;Y)$ are positive.

We may term $I_\alpha(X;Y)$ and $I_\alpha(Y;X)$ as α-informations of type I and type II.

In the limit, as $\alpha \to 1$, $I_\alpha(X;Y)$ and $I_\alpha(Y;X)$ tend respectively to $I_1(X;Y)$, information conveyed about X by Y, and $I_1(Y;X)$, information conveyed about Y by X, Shannon (1948). But $I_1(X;Y) = I_1(Y;X)$, so, as $\alpha \to 1$, both $I_\alpha(X;Y)$ and $I_\alpha(Y;X)$ tend to $I_1(X;Y)$, Shannon's definition of transformation.

2.2 α- Capacities

Two types of α-informations, defined above, in a natural way give rise to two types of α-capacities of type I and type II. α- capacity of type I may be defined as

$$C_{\alpha}^{(1)} = \max_{p(x_i)} I_{\alpha}(X;Y) = \max_{p(x_i)} \left[H_{\alpha}(X) - H_{\alpha}(X/Y) \right] \tag{3}$$

and α- capacity of type II may be defined as

$$C_{\alpha}^{(2)} = \max_{p(x_i)} I_{\alpha}(Y;X) = \max_{p(x_i)} \left[H_{\alpha}(Y) - H_{\alpha}(Y/X) \right] \tag{4}$$

Both (3) and (4), as indicated, are to be maximised over the input probabilities. In the limit, as $\alpha \to 1$, both tend to the Shannon's definition of capacity. They may be interpreted as the maximum of α-informations processed by the channel as perceived from the sending end and the receiving end respectively.

2.3 α-Redundancies And α-Efficiencies.

Two types of α-informations and α-capacities would give two types of α-redundancies, $\alpha > 0$, $\alpha \neq 1$, both absolute and relative, and two types of α-efficiencies $\alpha > 0$, $\alpha \neq 1$.

(i) Absolute α-Redundancies

Absolute α-redundancy of type I may be defined as

$$(\text{Abs.Redundancy})_{\alpha}^{1} = \text{Max } I_{\alpha}(X;Y) - I_{\alpha}(X;Y) \tag{5}$$

and absolute α-redundancy of type II as

$$(\text{Abs. Redundancy})^{2} = \text{Max } I_{\alpha}(Y;X) - I_{\alpha}(Y;X) \tag{6}$$

In the limit, as $\alpha \to 1$, both (5) and (6) tend to the known definition of absolute redundancy.

(ii) Relative α-Redundancies

Relative α-redundancy of type I may be defined as

$$(\text{Rel. Redundancy})_{\alpha}^{1} = \frac{(\text{Abs. Redundancy})_{\alpha}^{1}}{\text{Max } I_{\alpha}(X;\, Y)} \tag{7}$$

and relative α-redundancy of type II as

$$(\text{Rel. Redundancy})_{\alpha}^{2} = \frac{(\text{Abs. Redundancy})_{\alpha}^{2}}{\text{Max } I_{\alpha}(Y;X)} \tag{8}$$

In the limit, as $\alpha \to 1$, both (7) and (8) tend to the known definition of relative redundancy.

(iii) α-Efficiencies

α- efficiency of type I may be defined as

$$(\text{Efficiency})_{\alpha}^{1} = \frac{I_{\alpha}(X;\, Y)}{\text{Max } I_{\alpha}(X;Y)} \tag{9}$$

and α-efficiency of type II as

$$(\text{Efficiency})_{\alpha}^{2} = \frac{I_{\alpha}(Y;X)}{\text{Max } I_{\alpha}(Y;X)} \tag{10}$$

In the limit, as $\alpha \to 1$, both (9) and (10) tend to the known definition of efficiency.

(iv) Relation Between α-Efficiencies And α-Redundancies

By (5) ,(7) and (9) we have

and by (6), (8) and (10) we have

$$(\text{Rel.Redundancy})^2_\alpha = 1 - (\text{Efficiency})^2_\alpha$$

3. α-INFORMATION AND ADDITIVE PROPERTY

The situation when an input may be received by a number of outputs can arise (i) when the input symbols of a noisy channel are repeated a number of times rather than transmitted just once (ii) when the response to a single input is a sequence of output symbols rather than a single output symbol.

Suppose for each input symbol x_i from the input alphabet $X = \{x_i\}$ $i = 1, \ldots, n$ we receive two symbols y_{1j} and y_{2k}, y_{1j} from the output alphabet $Y_1 = \{y_{1j}\}$ $j = 1, \ldots, m$ and y_{2k} from the output alphabet $Y_2 = \{y_{2k}\}$ $k = 1, \ldots, r$. Without any loss of generality we may suppose that the two output symbols are received in the order y_{1j}, y_{2k}. Upon the reception of the first output symbol the 'a priori' probabilities $p(x_i)$ of the input symbols change into the 'a posteriori' probabilities $p(x_i/y_{1j})$ and upon the reception of the second output symbol $p(x_i)$ change into the 'even a more posteriori' probabilities $p(x_i / y_{1j}, y_{2k})$. So when two symbols y_{1j} and y_{2k} are received the α-entropy of the input symbols change from

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i) \qquad \alpha > 0, \alpha \neq 1$$

to the 'a posteriori' α - entropy

$$H_\alpha(X/y_{1j}) = \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i/y_{1j}) \tag{11}$$

$$\alpha > 0, \alpha \neq 1$$

and then to the 'even more a posteriori' α-entropy

$$H_\alpha(X/y_{1j}, y_{2k}) = \frac{1}{1-\alpha} \log \sum_i p^\alpha(x_i/y_{1j}, y_{2k}) \quad \alpha > 0, \alpha \neq 1 \qquad \ldots \quad (12)$$

So averaging (11) over the y_{1j} we shall have α-equivocation of X w.r.t. Y_1

$$H_\alpha(X/Y_1) = \sum_j p(y_{1j})\, H_\alpha(X/\, y_{1j}) \qquad (13)$$

and similarly on averging (12) over all the pairs (y_{1j}, y_{2k}) we shall have α-equivocation of X w.r.t. Y_1 and Y_2

$$H_\alpha(X/Y_1, Y_2) = \sum_{j,k} p(y_{1j}, y_{2k})\, H_\alpha(X/y_{1j}, y_{2k}) \qquad (14)$$

(13) and (14) suggest two ways of measuring α-information of type I i.e. information conveyed about X by Y_1 and Y_2 :

(i) We might define α-information conveyed about X by (Y_1, Y_2) just as we did when the channel output consisted of a single symbol i.e.

$$I_\alpha(X; Y_1, Y_2) = H_\alpha(X) - H_\alpha(X/Y_1, Y_2) \qquad \alpha > 0,\ \alpha \neq 1 \qquad (15)$$

(ii) We might consider the amount of α-information provided about X by Y_1 alone and then the amount of α-information conveyed about X after we have seen Y_1; this leads to considering

$$H_\alpha(X) - H_\alpha(X/Y_1) \qquad (16)$$

and

$$H_\alpha(X/Y_1) - H_\alpha(X/Y_1, Y_2) \qquad (17)$$

(16) has already been defined as $I_\alpha(X; Y_1)$ [by (1)] . It is

$$I_\alpha(X;Y_2/Y_1) = H_\alpha(X/Y_1) - H_\alpha(X/Y_1, Y_2) \quad \alpha > 0, \alpha \neq 1 \qquad (18)$$

and we may call this as α-information conveyed about X by Y_2 after we have observed Y_1.

By (77) of Chapter 2, both (15) and (18) are positive and in the limit, as $\alpha \to 1$, both tend to the corresponding concepts of mutual information dealt with in Abramson (1963).

Adding (16) and (18) and using (15) we have

$$I_\alpha(X;Y_1,Y_2) = I_\alpha(X;Y_1) + I_\alpha(X;Y_2/Y_1) \quad \alpha > 0, \alpha \neq 1 \qquad (19)$$

(19) expresses the additive property of α-information as a generalisation of additive property of mutual information since in the limit, as $\alpha \to 1$, it tends to the corresponding known result given on page 125, Abramson (1963). So combining this known result and our generalisation we can formally state it as

<u>Th. 1</u> The average α-information $\alpha > 0$ of type I conveyed about an input by a compound observation does not depend on whether we consider the compound observation as a whole or broken into its components.

The proof of the theorem when a compound observation consists of more than two elementary observations follows from (19) by induction. Suppose that the theorem is true for $m < n$ elementary observations constituting a compound observation, so when $n = m + 1$, by (19) we have

$$I_\alpha(X;Y_1,\ldots, Y_m,Y_{m+1}) = I_\alpha(X;Y_1,\ldots,Y_m) + I_\alpha(X;Y_{m+1}/Y_1,\ldots,Y_m) \qquad (20)$$

But by supposition

$$I_\alpha(X;Y_1,\dots,Y_m) = I_\alpha(X;Y_1) + I_\alpha(X;Y_2/Y_1) + \dots + I_\alpha(X;Y_m/Y_1,\dots,Y_{m-1})$$

and this reduces (20) to the result required.

4. α-INFORMATION OF SEVERAL ALPHABETS

The additive property of α-information of type I given in section 2 makes us consider a sequence of α-entropies, $\alpha > 0$, $\alpha \neq 1$

$$\begin{array}{l} H_\alpha(X) \\ H_\alpha(X/Y_1) \\ H_\alpha(X/Y_1,Y_2) \\ \dots\dots\dots \\ \dots\dots\dots \end{array}$$

Each member of this sequence is no greater than the proceeding one [by (77) of chapter 2] and the difference of any two successive members can be regarded as the α-information conveyed about X provided by a new observation :

$$\begin{aligned} I_\alpha(X;Y_1) &= H_\alpha(X) - H_\alpha(X/Y_1) \\ I_\alpha(X;Y_2/Y_1) &= H_\alpha(X/Y_1) - H_\alpha(X/Y_1,Y_2) \\ &\dots\dots\dots\dots\dots\dots \end{aligned}$$

$I_\alpha(X;Y_1)$ is the α-information conveyed about X by Y_1, $I_\alpha(X;Y_2/Y_1)$ is the α-information conveyed about X by Y_2 after we have observed Y_1 etc.

Following McGill (1954) we can define α-information $\alpha > 0$, $\alpha \neq 1$, of more than two alphabets. For three alphabets X, Y_1, Y_2 we may define

$$I_\alpha(X;Y_1;Y_2) = I_\alpha(X;Y_1) - I_\alpha(X;Y_1/Y_2) = H_\alpha(X) - H_\alpha(X/Y_1) - H_\alpha(X/Y_2) + H_\alpha(X/Y_1,Y_2) \qquad (21)$$

In the limit, as $\alpha \to 1$, (21) tends to a known result given in Abramson (1963).

It may be noted that for $\alpha > 0$, $\alpha \neq 1$, $I_\alpha(X;Y_1,Y_2)$, unlike $I_1(X;Y_1;Y_2)$, is not symmetrical in its arguments but, like $I_1(X;Y_1;Y_2)$ cannot be positive in general.

The above notion can be similarly extended to more than three alphabets.

5. CASCADED CHANNELS

We assume that a channel with its input symbols $X = \{x_i\}$ $i = 1,\ldots,n$ and output symbols $Y = \{y_j\}$ $j = 1,\ldots,m$ is cascaded with a second channel as in the figure below

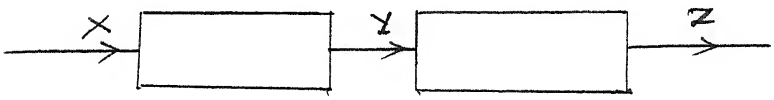

Fig. 1

with its input symbols identified with Y and output symbols consisting of $Z = \{z_k\}$ $k = 1,\ldots,t$.

These cascades give certain relationships among symbol probabilities. When x_i is transmitted, let the output of the first channel be y_j and let y_j produce z_k from the second channel. The symbol z_k depends on the original symbol x_i only through y_j, so the knowledge of intermediate symbol y_j makes the probability of obtaining the z_k depend only on y_j and not on the initial symbol x_i which produces y_j. This relationship may be written as

$$p(z_k/y_j,x_i) = p(z_k/y_j) \text{ for all } i,\ j,\ k \tag{22}$$

Since the probabilities $p(x_i/y_j)$ are known as 'a priori'

probabilities and (22) makes $p(z_k/y_j, x_i)$ also known, so the 'a posteriori' probabilities $p(x_i/y_j, z_k)$ for i,j,k are by Bayes' rule :

$$p(x_i/y_j, z_k) = \frac{p(x_i/y_j)\, p(z_k/x_i, y_j)}{\sum_i p(x_i/y_j)\, p(z_k/x_i, y_j)}$$

$$= \frac{p(x_i/y_j)\, p(z_k/y_j)}{\sum_j p(x_i/y_j) p(z_k/y_j)} \qquad \text{[by (22)]}$$

$$= p(x_i/y_j) \tag{23}$$

The result of leakage of information through such cascades of channels is generalised by proving that this is also valid for our concept of α-information of type I given by (1) i.e.

$$I_\alpha(X;Y) \geqslant I_\alpha(X;Z) \quad \alpha > 0, \quad \alpha \neq 1 \tag{24}$$

The proof of this follows from the following:

<u>Lemma</u> : As α- information is transmitted through cascaded channels from X to Y to Z, the α-equivocation increases i.e.

$$H_\alpha(X/Z) \geqslant H_\alpha(X/Y) \qquad \alpha > 0, \ \alpha \neq 1 \tag{25}$$

<u>Proof</u>. $H_\alpha(X/Z) - H_\alpha(X/Y) = H_\alpha(X/Z) - \frac{1}{1-\alpha} \sum_j p(y_j) \log \sum_i p^\alpha(x_i/y_j)$

$$= H_\alpha(X/Z) - \frac{1}{1-\alpha} \sum_{j,k} p(y_j, z_k) \log \sum_i p^\alpha(x_i/y_j)$$

$$= H_\alpha(X/Z) - \frac{1}{1-\alpha} \sum_{j,k} p(y_j, z_k) \log \sum_i p^\alpha(x_i/y_j, z_k)$$

[by (23)]

$$= H_\alpha(X/Z) - H_\alpha(X/Y, Z) \geq 0$$

[by (77) of chapter 2]

Q.E.D.

Proof of (24): Using (25)

$$I_\alpha(X;Y) = H_\alpha(X) - H_\alpha(X/Y) \geq H_\alpha(X) - H_\alpha(X/Z) = I_\alpha(X;Z)$$

Q.E.D.

For $\alpha = 1$ the result (24) is proved on page 115, Abramson (1963). The equality in (25) and hence in (24) will occur under the same conditions as for $\alpha = 1$; this case of equality for $\alpha = 1$ is also dealt with in the same reference.

6. BEHAVIOUR AND α-CAPACITIES OF SPECIFIC CHANNELS

6.1 Noiseless Channels

We know that in such channels every letter of the input alphabet is in 1-1 correspondence with a letter of the output alphabet. So joint probability matrix is of the form

$$P(X,Y) = \begin{bmatrix} p(x_1, y_1) & 0 \text{------} 0 \\ 0 & p(x_2, y_2) \ldots . 0 \\ 0 & 0 \text{------} p(x_n, y_n) \end{bmatrix}$$

and channel probability matrix and conditional probability matrices are

$$P(Y/X) = P(X/Y) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

(i) α- Entropies

$$H_\alpha(X,Y) = H_\alpha(X) = H_\alpha(Y) = (1/1-\alpha) \log \sum_i p^\alpha(x_i, y_i)$$

$$H_\alpha(Y/X) = H_\alpha(X/Y) = 0$$

(ii) α- Capacities

$$C_\alpha^{(1)} = \max \left[H_\alpha(X) - H_\alpha(X/Y)\right] = \operatorname{Max} H_\alpha(X) = \log n$$

$$C_\alpha^{(2)} = \max \left[H_\alpha(Y) - H_\alpha(Y/X)\right] = \log n$$

(iii) α- Redundancies

$$(\text{Abs.Redundancy})_\alpha^1 = \log n - H_\alpha(X)$$

$$(\text{Abs.Redundancy})_\alpha^2 = \log n - H_\alpha(Y) = (\text{Abs.Redundancy})_\alpha^1$$

$$(\text{Rel. Redundancy})_\alpha^1 = 1 - (H_\alpha(X)/\log n)$$

$$(\text{Rel.Redundancy})_\alpha^2 = 1 - (H_\alpha(Y)/\log n) = (\text{Rel.Redundancy})_\alpha^1$$

(iv) <u>α-Efficiencies</u>

$$(\text{Efficiency})_\alpha^1 = (H_\alpha(X)/\log n)$$

$$(\text{Efficiency})_\alpha^2 = (H_\alpha(Y)/\log n) = (\text{Efficiency})_\alpha^1$$

In fact we have

$$(\text{Efficiency})_\alpha^1 = (\text{Efficiency})_\alpha^2 = 1 - (\text{Rel.Redundancy})_\alpha^1$$

$$= (1 - \text{Rel.Redundancy})_\alpha^2$$

(iv) <u>Characteristics</u>

i) The α-uncertainties both at the sending and receiving ends are the same.

ii) Both the conditional α-entropy and α-equivocation are zero. This emphasizes a nonambiguous transmission irrespective of measuring information by any of the α-entropies $\alpha > 0$.

iii) Both the types of α-capacities, α-redundancies, absolute and relative, and α-efficiencies are same.

6.2. <u>CHANNELS WITH INDEPENDENT INPUT-OUTPUT</u>

It will be shown that such a system does not transmit α-information of either type, $\alpha > 0$, i.e. $I_\alpha(X;Y) = I_\alpha(Y;X) = 0$ and hence α- capacities of both types are zero.

We know that for such channels an input letter x_i can be received as any one of the letters y_j of the receiving alphabet with equal probability. So

$$P(X,Y) = \begin{array}{c|c} {}_{X}\diagdown^{Y} & \\ & \begin{bmatrix} p_1 & p_1 \cdots\cdots\cdots p_1 \\ p_2 & p_2 \cdots\cdots\cdots p_2 \\ \cdots & \cdots\cdots\cdots\cdots \\ p_n & p_n \cdots\cdots\cdots p_n \end{bmatrix} \end{array}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) = 1 \Rightarrow \sum_{i=1}^{n} m\, p_i = 1 \Rightarrow \sum_{i=1}^{n} p_i = \frac{1}{m}$$

$$p(x_i, y_j) = p_{ij} = p_i$$

$$p(x_i) = m\, p_i \quad \text{and} \quad p(y_j) = \frac{1}{m}$$

$$p(x_i)\, p(y_j) = p_i = p(x_i, y_j)$$

$$p(x_i/y_j) = mp_i \quad \text{and} \quad p(y_j/x_i) = 1/m$$

(i) <u>α - Entropies</u>

$$H_\alpha(X,Y) = \frac{1}{1-\alpha} \log \sum_{i,j} p^\alpha(x_i, y_j) = \frac{1}{1-\alpha} (\log m + \log \sum_i p_i^\alpha) \quad \ldots (26)$$

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_i (m\, p_i)^\alpha = \frac{\alpha}{1-\alpha} \log m + \frac{1}{1-\alpha} \log \sum_i p_i^\alpha \quad (27)$$

$$H_\alpha(Y) = \frac{1}{1-\alpha} \log \sum_j (\sum_i p_i)^\alpha = \frac{1}{1-\alpha} \log \sum_j (1/m)^\alpha = \log m \quad (28)$$

$$H_\alpha(X/Y) = \frac{1}{1-\alpha} \sum_i p(y_j) \log \sum_i p^\alpha(x_i/y_j) = \frac{1}{1-\alpha} \log \sum_i m^\alpha p_i^\alpha$$

$$= H_\alpha(X) \quad \ldots (29)$$

[by (27)]

$$H_\alpha(Y/X) = \frac{1}{1-\alpha} \sum_i p(x_i) \log \sum_j p^\alpha(y_j/x_i) = \log m = H_\alpha(Y) \quad (30)$$

[by (28)]

The statement made in the beginning of this section follows from ~~by~~ (29) and (30).

Hence we can conclude that if we use α-entropies to study such channels, they retain their special feature of having the largest internal loss in contrast to noise-free channels which have the least.

6.3 CHANNELS WITH SYMMETRIC NOISE STRUCTURES

It will be shown that for such channels the conditional α-entropy $\alpha > 0$ is independent of input probabilities and hence $C_\alpha^{(2)}$ is the same for all $\alpha > 0$ but, unlike as for $\alpha = 1$, $I_\alpha(X;Y)$, for $\alpha > 0$, $\alpha \neq 1$, is not maximum for equiprobable x_i even when all the $p(y_j)$ are equal.

Since in such channels each input letter in transformed into a finite number of output letters with a similar set of probabilities for all the input probabilities so the channel matrix $P(Y/X)$ contains identical rows and identical columns but not necessarily in the same position. Consequently conditional α-entropy pertinent to the letter x_i will be

$$H_\alpha(Y/x_i) = \frac{1}{1-\alpha} \log \sum_{j=1}^{m} p^\alpha(y_j/x_i) = \text{constant} = h \text{ (say)}$$

for all x_i

Hence

$$H_\alpha(Y/X) = \sum_{i=1}^{n} p(x_i)\, H_\alpha(Y/x_i) = h \qquad \alpha > 0,\ \alpha \neq 1 .$$

The result is known to be true for $\alpha = 1$.

$$C_\alpha^{(2)} = \text{Max}\left[H_\alpha(Y) - H_\alpha(Y/X)\right] = \log m - h$$

If $I_\alpha(Y;X) = C_\alpha^{(2)}$ then all $p(y_j)$ are equal and this implies that in this case all $p(x_i)$ will also be equal and matrix $P(X/Y)$ will also have identical rows. So in this case $I_\alpha(X;Y) = \log n - h'$ where $h' = H_\alpha(X/y_j)$ for all y_j. But in this case $\log n - h'$ will not be in general equal to $C_\alpha^{(1)}$. The following example would show this :

Let the channel matrix be

$$\begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix} \quad \text{and} \quad p(x_1) = 11/20 \text{ , } p(x_2) = 9/20 \text{ .}$$

Considering the base of the logs to be 10 we shall have

$$H_{1/3}(X) = 1.5 \log \left[(11/20)^{1/3} + (9/20)^{1/3} \right] = .30045 \quad (31)$$

$$H_{1/3}(X/Y) = 1.5 \quad (31/60) \left[\log \left((22/31)^{1/3} + (9/31)^{1/3} \right) + (29/60) \log \left((11/29)^{1/3} + (18/29)^{1/3} \right) \right]$$

$$= .29172 \quad (32)$$

When $p(x_1) = p(x_2) = 1/2$ we shall have

$$H_{1/3}(X) = .3010 \quad (33)$$

$$H_{1/3}(X/Y) = 1.5 \log \left[(2/3)^{1/3} + (1/3)^{1/3} \right] = .29265 \quad (34)$$

From (31) and (32)

$$H_{1/3}(X) - H_{1/3}(X/Y) = .00873 \quad (35)$$

and from (33) and (34)

$$H_{1/3}(X) - H_{1/3}(X/Y) = .00835 \tag{36}$$

The result follows from (35) and (36)

6.4 B S C

Let $p(0) = p(x_1) = r$; $p(1) = p(x_2) = 1 - r$

$p(1/1) = p(0/0) = p$; $p(0/1) = p(1/0) = q$

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left[r^\alpha + (1-r)^\alpha\right]$$

$$H_\alpha(Y/x_1) = \frac{1}{1-\alpha} \log (p^\alpha + q^\alpha)$$

$$H_\alpha(Y/X) = \sum_i p(x_i) \left[\frac{1}{1-\alpha} \log (p^\alpha + q^\alpha)\right]$$

$$= \frac{1}{1-\alpha} \log (p^\alpha + q^\alpha)$$

Since by section 2.3 of chapter 2 $H_\alpha(X)$ is maximum when all the $p(x_i)$ are equal, so by (4)

$$C_\alpha^{(2)} = \max \left[H_\alpha(Y) - \frac{1}{1-\alpha} \log (p^\alpha + q^\alpha)\right] = 1 - \frac{1}{1-\alpha} \log(p^\alpha + q^\alpha) \tag{37}$$

$$\alpha > 0, \ \alpha \neq 1$$

$$C_\alpha^{(1)} = \operatorname{Max}\left[\frac{1}{1-\alpha}\left\{\log\left(p^\alpha(x_1)+p^\alpha(x_2)\right) - p(y_1)\log\left(p^\alpha(x_1/y_1)+p^\alpha(x_2/y_1)\right) - p(y_2)\log\left(p^\alpha(x_1/y_2)+p^\alpha(x_2/y_2)\right)\right\}\right]$$

$$= \operatorname{Max}\left[\frac{1}{1-\alpha}\left\{p(y_1)\log\frac{p^{\alpha}(x_1)+p^{\alpha}(x_2)}{p^{\alpha}(x_1/y_1)+p^{\alpha}(x_2/y_1)}+p(y_2)\log\frac{p^{\alpha}(x_1)+p^{\alpha}(x_2)}{p^{\alpha}(\frac{x_1}{y_2})+p^{\alpha}(\frac{x_2}{y_2})}\right\}\right]$$

$$= \operatorname{Max}\left[\frac{\alpha}{1-\alpha}(p\,p(x_1)+qp(x_2))\log\frac{\left(p^{\alpha}(x_1)+p^{\alpha}(x_2)\right)(p\,p(x_1)+q\,p(x_2))}{p^{\alpha}\,p^{\alpha}(x_1)+q^{\alpha}\,p^{\alpha}(x_2)}\right.$$

$$\left.+\left(qp(x_1)+pp(x_2)\right)\log\frac{\left(p^{\alpha}(x_1)+p^{\alpha}(x_2)\right)(q\,p(x_1)+pp(x_2))}{q^{\alpha}\,p^{\alpha}(x_1)+p^{\alpha}p^{\alpha}(x_2)}\right] \quad (38)$$

To my knowledge no method to maximise (38) subject to $p(x_1) + p(x_2) = 1$, $p(x_i) > 0 \quad i = 1,2$ so far exists.

In passing we may note that for such channels

$$C_{\alpha}^{(2)} \leq C_1 \leq C_{\beta}^{(2)} \quad 0 < \alpha < 1 \text{ and } \beta > 1$$

since

$$H_{\alpha}(p,\ 1-p) \geqslant H_1(p,\ 1-p) \geqslant H_{\beta}(p,\ 1-p)$$

6.5 <u>B E C</u>

As already known that for such channels input contains two symbols $\{0,\ 1\}$ and out-put three symbols $\{0,\ y,\ 1\}$. The letter y indicates that the output is erased and no deterministic decision can be made as to whether the transmitted letter was 0 or 1.

Letting $p(0) = r$, $p(1) = 1-r$; $p(0/0) = p(1/1) = p$

$p(y/1) = p(y/0) = q$; $p(0/1) = p(1/0) = 0$ and $p + q = 1$

We shall have

$$H_\alpha(X/Y) = q\ H_\alpha(X)$$

$$I_\alpha(X;Y) = H_\alpha(X) - (1-p)\ H_\alpha(X) = p\ H_\alpha(X)$$

Hence

$$C_\alpha^{(1)} = p \operatorname{Max} H_\alpha(X) = p \tag{39}$$

$$H_\alpha(Y) = (1/1-\alpha)\log\left((r\ p)^\alpha + q^\alpha + (1-r)^\alpha\ p^\alpha\right)$$

$$= (1/1-\alpha)\log\left[p^\alpha(r^\alpha + (1-r)^\alpha) + q^\alpha\right]$$

$$H_\alpha(Y/X) = (1/1-\alpha)\log(p^\alpha + q^\alpha)$$

Hence

$$C_\alpha^{(2)} = \operatorname*{Max}_{\{r,1-r\}}\left[(1/1-\alpha)\log\{(r\ p)^\alpha + q^\alpha + (1-r)^\alpha p^\alpha\}\right] - (1/1-\alpha)\log(p^\alpha + q^\alpha)$$

I think that the expression $(1/1-\alpha)\log((rp)^\alpha + q^\alpha + (1-r)^\alpha\ p^\alpha)$ can be maximised by the technique of geometric programming, Duffin, Peterson and Zener (1967).

7. VERIFICATIONS OF RESULTS FOR ALTERNATIVE DEFINITION OF CONDITIONAL α-ENTROPY.

In section 5 of chapter 2 are developed alternative definitions of conditional α-entropy and α-equivocation. Rewriting them we have

$$H_\alpha(Y/X) = (1/1-\alpha)\log \sum_i p(x_i) \sum_j p^\alpha(y_j/x_i) \qquad \alpha > 0,\ \alpha \neq 1 \tag{40}$$

$$H_\alpha(X/Y) = (1/1-\alpha)\log \sum_j p(y_j) \sum_i p^\alpha(x_i/y_j) \qquad \alpha > 0,\ \alpha \neq 1 \tag{41}$$

There it was also shown that for $0<\alpha<1$ $H_\alpha(Y/X)$ given by (5) of chapter 2 is less than that given by (40) and for $\alpha > 1$ it is more than (40). Similar interpretation can be had for $H_\alpha(X/Y)$: Hence α-informations defined by (1) and (2) would be less for $0<\alpha<1$ and more for $\alpha > 1$ if we measure conditional α-entropy and α-equivocation by (40) and (41) in place of (5) and (6) of chapter 2.

All the results, except those given by (32), (38) and (39), proved in this chapter are also true for these alternative-definitions. If we define $H_\alpha(Y/X)$ and $H_\alpha(X/Y)$ by (40) and (41), (32) would become

$$H_{1/3}(X/Y) = .29175$$

and the result proved there is still valid. (38) would become

$$C_\alpha^{(1)} = \max\left[\frac{1}{1-\alpha}\log\left\{\frac{p^\alpha(x_1)+p^\alpha(x_2)}{\big(pp(x_1)+qp(x_2)\big)^{1-\alpha}\left(p^\alpha p^\alpha(x_1)+q^\alpha p^\alpha(x_2)\right)+\big(qp(x_1)+pp(x_2)\big)^{1-\alpha}\left(q^\alpha p^\alpha(x_1)+p^\alpha p^\alpha(x_2)\right)}\right\}\right]$$

$$\alpha > 0,\ \alpha \neq 1 \qquad \ldots.(42)$$

But the evaluation of (42) is as difficult as that of (25).

(39) would become

$$C_\alpha^{(1)} = 1 - \max\left[\frac{1}{1-\alpha}\log p + q\left(r^\alpha + (1-r)^\alpha\right)\right]$$

$$\alpha > 0,\ \alpha \neq 1$$

The expression in bracket too, I think can be maximised by the technique of geometric programming.

The above three violations involve computations. So on the whole all the main theoretic results proved in this chapter are true for the alternative definitions given by (40) and (41).

CHAPTER - 4

SHANNON'S FUNDAMENTAL CODING THEOREMS

1. INTRODUCTION

Shannon (1948) gave two coding theorems for discrete memoryless channels. These theorems led to the development of a very useful application of information theory viz. The Coding Theory. In this chapter we shall generalise these theorems with the help of our concepts of α-informations and α-capacities. Since these generalisations will tend in the limit, as $\alpha \to 1$, to the present available statements of these theorems, so in a way our efforts would be directed to embed these theorems with a dynamic character.

In section 2 a generalisation of Shannon's fundamental theorem for noisy discrete memoryless channels is stated and the proof is given for binary symmetric channels. In section 3 Fano bound is generalised and on its basis the converse of Shannon's fundamental theorem for noisy discrete memoryless channels is generalised for a family of extensions of such a channel. In section 4 a generalisation, based on the generalisation of Shannon's theorem for noiseless coding given in Abramson (1963), of Campbell (1965) coding theorem is given. Since Campbell coding theorem is a generalisation of Shannon's theorem and our generalisation is a generalisation of Campbell's theorem so our generalisation is also a generalisation of Shannon's theorem. In section 5 taking the usual average length $\bar{L}=\sum_{i=1}^{n} p(x_i)n_i$ associated with the transmission of a set of uniquely decipherable code words of lengths $[n_1, n_2, \ldots, n_n]$ with

probabilities $[p(x_1), \ldots, p(x_n)]$, α- bounds for $\bar{L}$ in terms of Renyi's entropy are found. As consequences of these α-bounds the requirement of successive equiprobable partitioning for the optimality of Shannon-Fano code and the sufficient condition of the known result for $\bar{L}$ attaining the lower bound when entropy is measured by Shannon's definition are proved. Some problems that these bounds give birth to are also given.

2. A GENERALISATION OF SHANNON'S SECOND FUNDAMENTAL THEOREM

By (2) of chapter 3 $[H_\alpha(Y) - H_\alpha(Y/X)]$ is the transferred α-information of type II and may be termed as rate of order α (in abbreviation ''α-rate'') $\alpha > 0$, $\alpha \neq 1$ at which α-information of type II per symbol is transferred through the channel. Since the base of logarithm has thro'out been taken to be 2, so ''bit'' may again be taken as the unit of its measure. The maximum of this α-rate has been defined as α-capacity of type II.

Similar interpretation may be had for the expression $[H_\alpha(X) - H_\alpha(X/Y)]$.

<u>Statement</u>

It is possible to find an appropriate encoding procedure to encode the output of S, a discrete memoryless source, so that the encoded output can be transmitted through a noisy channel at an α-rate $[H_\alpha(Y) - H_\alpha(Y/X)] < C_\alpha^{(2)}$, $\alpha > 0$, $\alpha \neq 1$ (α is fixed for a particular consideration) and decoded with as small a probability of error as desired.

In the limit, as $\alpha \to 1$, the above statement tends to an already available statement.

A Heuristic Proof For BSC

Let $\{X\} = [x_1, x_2, \ldots\ldots, x_N]$ be a message ensemble of the source S

Assumptions

(i) the N messages are selected with equal probability i.e., the selection is made completely at random;

(ii) the channel is specified by

$$p(0/0) = p(1/1) = p,\ p(0/1) = p(1/0) = q,\ q < 1/2 \text{ and } p+q = 1$$

(iii) all the encoded messages have the same length of n binary digits;

(iv) the binary endoding procedure is a random one.

Decoding procedure

We have N source messages each encoded with n binary digits. At the receiver we have a catalog containing all possible 2^n n-symbol sequences with N source sequences specifically underlined. In case the received message does not agree with any of the N source sequences, we take that source sequence as being transmitted which differs from the received sequence in least no. of digits.

A sequence of n 1's and 0's can be considered as a sequence of n independent Bernoulli trials in which the correct transmission of each binary digit occurs with probability p and erroneous transmission with probability $q = 1 - p < 1/2$. Taking Z as a random variable to represent

the no. of erroneous digits in a received message, we have

$$P(Z = k) = \binom{n}{n-k} p^{n-k} q^{k}$$

so

$$E(Z) = \sum_{k=0}^{n} k\binom{n}{n-k} p^{n-k} q^{k} = nq$$

Thus we expect, on the average, in a sequence nq digits to be altered by noise. Therefore, according to our decoding procedure, from the list of underlined sequences representing N source sequences at the receiving end each sequence that differs from the received sequence by nq or less digits could be expected to be the transmitted sequence. So the expected no. of sequences that can be considered as possible original message for a transmitted sequence is

$$M = \sum_{k=0}^{nq} \binom{n}{k} \tag{1}$$

For $q < 1/2$, (1) can be written as

$$M \leq 1 + nq \binom{n}{nq} \tag{2}$$

But for large n we have by Strling's formula

$$n! \approx \sqrt{2\pi}\, e^{-n}\, n^{n+\frac{1}{2}} \tag{3}$$

Using (3), (2) reduces to

$$M \leqslant 1 + \sqrt{(nq)/2\pi p} \quad p^{-np} \; q^{-np} \tag{4}$$

[For details of this step see Reza(1961)]

of these M sequences, only one is correct and M-1 are potential misinterpretations of the transmitted signal.

By assumption (iv) the probability of selecting an n-digit sequence so as to correspond it to one of the N messages of $\{X\}$ is $N/2^n$; this in turn reduces the expected no. of misinterpretations of the received sequence for a transmitted sequence corresponding to one of the N messages of $\{X\}$ from M-1 to $(N/2^n)(M-1)$. So denoting the frequency of the occurrence of an error by M_A we have

$$M_A \leqslant (N/2^n)\,(M-1) \tag{5}$$

Using (4), (5) can be written as

$$M_A \leqslant (N/2^n)\sqrt{(nq)/2\pi p} \quad p^{-np} \; q^{-nq} \tag{6}$$

We know by (37) of chapter 3 that for a BSC,

$$C_\alpha^{(2)} = 1 - (1/1-\alpha)\log_2(p^\alpha + q^\alpha)$$

$$\text{or} \quad \frac{1}{2} = (p^\alpha + q^\alpha)^{-1} \left(2^{-C_\alpha^{(2)} - \alpha + \alpha C_\alpha^{(2)}}\right)$$

$$\text{or} \quad (1/2)^n = (p^\alpha + q^\alpha)^{-n} \left(2^{-C_\alpha^{(2)} - \alpha + \alpha C_\alpha^{(2)}}\right)^n \tag{7}$$

Using (7), (6) can be written as

$$M_A \leqslant N \frac{2^{-n\, C_\alpha^{(2)}}\ 2^{\alpha n (C_\alpha^{(2)} - 1)}\ p^{-np}\ q^{-nq}}{(p^\alpha + q^\alpha)^n} \sqrt{\frac{nq}{2\pi\, p}} \tag{8}$$

<u>Case (i) when $0 < \alpha < 1$</u>

Now by suitably choosing N, the no. of messages of $\{X\}$ to be equal to or less than $2^{n\, C_\alpha^{(2)}} / n$, (8) reduces to

$$M_A \leqslant \left[\frac{p^{-p}\ q^{-q}}{(p^\alpha + q^\alpha)\ 2^{\alpha(1 - C_\alpha^{(2)})}} \right]^n \sqrt{\frac{q}{2n\pi\, p}} \tag{9}$$

Now we shall prove that

$$\frac{p^{-p}\ q^{-q}}{(p^\alpha + q^\alpha)\ 2^{\alpha(1 - C_\alpha^{(2)})}} \leqslant 1 \tag{10}$$

(10) can be written as

$$(p^\alpha + q^\alpha)\ 2^{\alpha(1 - C_\alpha^{(2)})} \geqslant (1/ p^p\ q^q)$$

$$\text{or} \quad \log (p^\alpha + q^\alpha) + \alpha(1 - C_\alpha^{(2)}) \geqslant - \log (p^p\ q^q)$$

$$\text{or} \quad (1-\alpha)\ H_\alpha(p,q) + \alpha(1 - C_\alpha^{(2)}) \geqslant H_1(p,q) \tag{11}$$

But from (37) of Chapter 3 we have, in particular for $0 < \alpha < 1$,

Now we shall prove

$$\frac{p^{-p} \quad q^{-q}}{(p^{\alpha}+q^{\alpha}) \left[C_{\alpha}^{(2)} - \alpha\,(C_{\alpha}^{(2)} -1)- \beta\right]} \leq 1 \tag{13}$$

2

(13) can be written as

$$\log(p^{\alpha}+q^{\alpha}) + C_{\alpha}^{(2)} - \alpha(C_{\alpha}^{(2)} - 1) - \beta \geqslant - \log(p^{p} q^{q})$$

or

$$(1-\alpha) (1- C_{\alpha}^{(2)})+ C_{\alpha}^{(2)} - \alpha(C_{\alpha}^{(2)} -1)- \beta \geqslant -\log(p^{p} q^{q})$$

or

$$1- C_{\alpha}^{(2)} - \alpha+ \alpha C_{\alpha}^{(2)}+ C_{\alpha}^{(2)} - \alpha C_{\alpha}^{(2)} +\alpha - \beta \geqslant 1 - C_1$$

or $\quad 1 - \beta \geqslant 1 - C_1$

or $\quad \beta \leqslant C_1$

Conversely by our supposition

$$\beta \leq C_1$$

or $\quad 1 - \beta \geqslant 1 - C_1$

and further retracing the steps we shall have (13).
So (13) is established, hence from (12) $M_A \to 0$ as $n \to \infty$ irrespective of the noise characteristic $(q < 1/2)$.

This completes the proof.

3. GENERALISATION OF FANO BOUND AND THE CONVERSE OF SHANNON'S SECOND FUNDAMENTAL THEOREM.

3.1 A Generalisation of Fano Bound

Fano bound gives upper bound for equivocation in terms of the probability of error [page 153, Abramson (1963)]. Below this bound will be generalised i.e., an upper bound, which tends to the known result, as $\alpha \to 1$, will be established for α- equivocation, $\alpha > 0$, $\alpha \neq 1$.

Suppose a decision rule $d(y_j)$ for a channel with input $\{X\} = [x_1, \ldots, x_n]$ and output $\{Y\} = [y_1, \ldots, y_m]$ be such that assigns to $d(y_j)$ that x_i (say) x^* for which

$$p(x^*/y_j) \geqslant p(x_i/y_j) \qquad \text{for all } i. \tag{14}$$

We know that this decision rule is sometimes called a conditional maximum likelihood decision rule. By Bayes' law, (14) can be written as

$$\frac{p(y_j/x^*)\, p(x^*)}{p(y_j)} \geqslant \frac{p(y_j/x_i)\, p(x_i)}{p(y_j)} \quad \text{for all } i .$$

When the a priori probabilities are all equal, the above rule can be written as

$$d(y_j) = x^*$$

subject to

$$p(y_j/x^*) \geqslant p(y_j/x_i) \quad \text{for all } i.$$

We know the decision rule so defined is known as the maximum likelihood decision rule.

The error probability P_E using any given decision rule is given by

$$P_E = \sum_j \left[p(y_j) \left\{ 1 - p(d(y_j)/y_j) \right\} \right] \qquad (15)$$

(15) can alternatively be written as

$$P_E = \sum_Y p(y) - \sum_Y p(d(y)/y)\, p(y)$$

$$= 1 - \sum_Y p(d(y),\, y)$$

Writing $\bar{P}_E = 1 - P_E$ we have

$$\bar{P}_E = \sum_Y p(x^*,\, y) \qquad (16)$$

and

$$P_E = \sum_{Y, X-x^*} p(x,y) \qquad (17)$$

<u>Case (i) when $0 < \alpha < 1$</u>

By (7) we have

$$H_\alpha(X/Y) = \frac{1}{1-\alpha} \sum_Y p(y) \;\log\; \sum_X p^\alpha(x/y)$$

But by Jensen's inequality

$$\sum_Y p(y) \log \sum_X p^\alpha(x/y) \lesssim \log\left[\sum_Y p(y) \left\{ \sum_{X-x^*} p^\alpha(x/y) + p^\alpha(x^*/y) \right\} \right]$$

$$= \log\left[\sum_{Y, X-x^*} p(y) p^\alpha(x/y) + \sum_Y p(y) p^\alpha(x^*/y) \right]$$

$$= \log\left[\sum_{Y, X-x^*} p^\alpha(y) p^\alpha(x/y) p^{1-\alpha}(y) + \sum_Y p^\alpha(y) p^\alpha(x^*/y) p^{1-\alpha}(y) \right]$$

$$\leqslant \log \left[\left\{ \sum_{Y, X-x^*} p(x,y) \right\}^{\alpha} \left\{ \sum_{Y, X-x^*} p(y) \right\}^{1-\alpha} + \left\{ \sum_{Y} p(x^*,y) \right\}^{\alpha} \left\{ \sum_{Y} p(y) \right\}^{1-\alpha} \right]$$

(By Hölder inequality)

$$= \log \left[\left\{ \sum_{Y, X-x^*} p(x,y) \right\}^{\alpha} (n-1)^{1-\alpha} + \left\{ \sum_{Y} p(x^*,y) \right\}^{\alpha} \right] \tag{18}$$

Multiplying (18) by $(1/1-\alpha)$, a positive number, and using (16) and (17) we have

$$H_{\alpha}(X/Y) \leqslant (1/1-\alpha) \log \left[P_E^{\alpha} (n-1)^{1-\alpha} + \bar{P}_E^{\alpha} \right] \tag{19}$$

$$0 < \alpha < 1$$

Case (ii) when $\alpha > 1$

Since $\sum_X p^{\alpha}(x/y) \leqslant 1$, so

$$\sum_Y p(y) \log \sum_X p^{\alpha}(x/y) \geqslant \sum_Y p(y) \sum_X p^{\alpha}(x/y) \log \sum_X p^{\alpha}(x/y)$$

$$\geqslant \left[\sum_Y p(y) \sum_X p^{\alpha}(x/y) \right] \log \left[\sum_Y p(y) \sum_X p^{\alpha}(x/y) \right] \tag{20}$$

(By Jensen's inequality)

But $\sum_Y p(y) \sum_X p^{\alpha}(x/y) \geqslant 1$, let it be β ; so using this fact

(20) reduces to

$$\sum_Y p(y)\log \sum_X p^{\alpha}(x/y) \geqslant \beta \log\left[\sum_Y p(y) \sum_X p^{\alpha}(x/y)\right]$$

$$= \beta\log\left[\sum_{Y,X-x^*} p(y)p^{\alpha}(x/y)+\sum_Y p^{\alpha}(x^*/y)p(y)\right]$$

$$= \beta\log\left[\sum_{Y,X-x^*} p^{\alpha}(x,y)p^{1-\alpha}(y)+ \sum_Y p^{\alpha}(x^*,y)p^{1-\alpha}(y)\right]$$

$$\geqslant \beta \log\left[\left\{\sum_{Y,X-x^*} p(x,y)\right\}^{\alpha}(n-1)^{1-\alpha}+\left\{\sum_Y p(x^*,y)\right\}^{\alpha}\right] \tag{21}$$

(By Holder inequality)

Multiplying (21) by $(1/1-\alpha)$, a negative number, using (16) and (17) and the fact that $\beta \leqslant 1$, we have

$$H_{\alpha}(X/Y) \leqslant (1/1-\alpha) \log\left[P_E^{\alpha}(n-1)^{1-\alpha}+\bar{P}_E^{\alpha}\right] \tag{22}$$

$$\alpha > 1,\ \alpha \neq 1$$

So combining (19) and (22) we have

$$H_{\alpha}(X/Y) \leqslant (1/1-\alpha) \log\left[P_E^{\alpha}(n-1)^{1-\alpha}+ \bar{P}_E^{\alpha}\right] \tag{23}$$

$$\alpha > 0,\ \alpha \neq 1.$$

In the limit, as $\alpha \to 1$, (23) tends to

$$H_1(X/Y) \leqslant H_1(P_E) + P_E \log(n-1)$$

as already known.

3.2 A Generalisation of The Converse Of Shannon's Second Fundamental Theorem

A form of the converse of Shannon's 2nd fundamental theorem as applicable to α-entropies is stated and the proof is given for a class of channels - a family of extensions of any discrete memoryless channel - and when the input probabilities are equal. For this we need the following lemma.

Lemma : α-capacity of either type of an nth order extension of a discrete memoryless channel is n times the α-capacity of the corresponding type of the channel.

We know that an nth order extension of a discrete memoryless channel with input $\{X\} = [x_1,\dots,x_n]$ and output $\{Y\} = [y_1,y_2,\dots,y_m]$ has all n^n n-tuples of the form $(x_1,x_2,\dots,x_n) = u$ (say) as its input and all m^n n-tuple of the form $(y_1,\dots,y_n) = v$ (say) as its output and

$$p(u) = p(x_1) \dots p(x_n)$$

$$p(v) = p(y_1) \dots p(y_n)$$

$$p(v/u) = p(y_1/x_1) \dots p(y_n/x_n)$$

$$p(u/v) = p(x_1/y_1) \dots p(x_n/y_n).$$

α- information of type I for such an extension is given

$$I_\alpha(U;\ V) = H_\alpha(U) - H_\alpha(U/V) \tag{24}$$

where U is the set of all n-tuples u and V the set of all n-tuples v. But

$$H_\alpha(U) = (1/1-\alpha) \log \sum_U p^\alpha(u)$$

$$= (1/1-\alpha)\log \sum_U p^\alpha(x_1).....p^\alpha(x_n)$$

$$= n\, H_\alpha(X) \qquad (25)$$

$$H_\alpha(U/V) = (1/1-\alpha) \sum_V p(v) \log \sum_U p^\alpha(u/v)$$

$$= (1/1-\alpha) \sum_V p(v) \log \sum_U p^\alpha(x_1/y_1)...p^\alpha(x_n/y_n)$$

$$=(n/1-\alpha) \sum_V p(v) \log \sum_X p^\alpha(x_i/y_j)$$

$$= n\, H_\alpha\,(X/Y) \qquad (26)$$

Using (25) and (26), (24) reduces to

$$I_\alpha(U;\ V) = n\, I_\alpha(X;\ Y) \qquad (27)$$

α- capacity of type I is given by

$$\operatorname*{Max.}_U I_\alpha(U;\ V)$$

so from this and (27) the statement of the lemma for this part follows.

Similarly we can prove that

$$I_\alpha(V;\ U) = n\, I_\alpha(Y;\ X) \qquad (28)$$

Since α-capacity of type II is given by

$$\underset{U}{\text{Max.}}\ I_{\alpha}(V;\ U)$$

so from this and (28) the result follows for α-capacity of type II.

<u>Statement</u>

Given $\varepsilon > 0$, it is impossible to encode a set of messages in $2^{n(C_{\alpha}^{(1)} + \varepsilon)}$ code words so as to be transmitted through an nth extension of a noisy discrete memoryless channel having α-capacity $C_{\alpha}^{(1)}$ with as small a probability of error as desired.

As $\alpha \to 1$, this tends to an already available statement for such a family of extensions.

<u>Proof:</u> Suppose the nth extension, of the type considered in the lemma, of the channel has $2^{n(C_{\alpha}^{(1)} + \varepsilon)} = N$ (say) code words, each to be transmitted with probability $1/N$. So by the above lemma and (23) for $0 < \alpha < 1$ we have

$$n\, C_{\alpha}^{(1)} \geqslant H_{\alpha}(U) - H_{\alpha}(U/V) \geqslant \log \frac{N}{\left\{ P_E^{\alpha}(N-1)^{1-\alpha} + \bar{P}_E^{\alpha} \right\}^{1/(1-\alpha)}}$$

$$\left(2^{nC_{\alpha}^{(1)}}/N\right) \geqslant \left(\frac{1}{P_E^{\alpha}(N-1)^{1-\alpha} + \bar{P}_E^{\alpha}} \right)^{(1/1-\alpha)}$$

$$\left(N/2^{n\,C_{\alpha}^{(1)}}\right)^{1-\alpha} \leqslant P_E^{\alpha}(N-1)^{1-\alpha} + \bar{P}_E^{\alpha} \qquad (29)$$

putting $N = 2^{n(C_\alpha^{(1)} + \epsilon)}$ to the left side of (29), it reduces to

$$(2^{n\epsilon})^{1-\alpha} \leq P_E^\alpha (N-1)^{1-\alpha} + (1 - P_E)^\alpha \qquad (30)$$

For any α from $0 < \alpha < 1$, the right side of (30) is 1 iff $P_E = 0$ otherwise it is always more than 1; but the left side of (30) is always more than 1 and cannot be made equal to 1 howsoever small n and ϵ we may take; actually for a fixed ϵ, the more the value of n we take, the more away will P_E be from 0.

Again by the above lemma and (23) for α 1, we have

$$\frac{2^{n C_\alpha^{(1)}}}{N} \geq \left[\frac{1}{P_E^\alpha (N-1)^{1-\alpha} + (1-P_E)^\alpha}\right]^{\frac{1}{1-\alpha}}$$

$$\left(\frac{2^{n C_\alpha^{(1)}}}{N}\right)^{\alpha-1} \geq P_E^\alpha (N-1)^{1-\alpha} + (1 - P_E)^\alpha$$

$$\left(\frac{1}{2^{n\epsilon}}\right)^{\alpha-1} \geq P_E^\alpha (N-1)^{1-\alpha} + (1 - P_E)^\alpha \qquad (31)$$

For any α, from $\alpha > 1$, the right side of (31) is 1 iff $P_E = 0$ otherwise it is always less than 1, but the left side of (31) is always less than 1 and the more the

n the smaller the left side and for the inequality (31) to hold the term $(1 - P_E)^{\alpha}$ on its right side should be made comparatively small and this requires P_E to become more and more i.e., to deviate away and away from 0.

This completes the proof.

4. A GENERALISATION OF SHANNON'S FIRST FUNDAMENTAL THEOREM

For a uniquely decipherable code with word length given by the sequence $\{n_i\}$ and transmitted through a noiseless channel with probabilities given respectively by the sequence $\{p(x_i)\}$, Campbell (1965) defined a cost function of order t, $0 < t < \infty$,

$$L(t) = (1/t) \log_D \sum_i p(x_i)\, D^{t\, n_i} \tag{32}$$

where $(1/1+t) = \alpha$ and D is the number of letters in the encoding alphabet, and proved a coding theorem for noiseless channels when the entropy of the source is measured by

$$H_\alpha(X) = (1/1-\alpha) \log_D \sum_i p^{\alpha}(x_i) \qquad \alpha > 0,\ \alpha \neq 1 \tag{33}$$

The results he obtained tend in the limit, as $\alpha \to 1$, to the already available results of the corresponding Shannon's first coding theorem for such channels.

A generalisation of Shannon's first theorem is given in Abramson (1963). There it has been argued that since for a received symbol y_j we can associate the a posteriori

probabilities $p(x_i/y_j)$ with the input alphabet, so keeping in view the interpretation of Shannon's first theorem for the entropy of an alphabet as the average number of D-ary letters necessary to represent one symbol of the alphabet, the average of the a posteriori entropy $H_1(X/y_j)$ over y_j may likewise be interpreted. This idea gave rise to that generalisation.

The difference between Shannon's first theorem and that generalisation lies in the fact that the former deals with a fixed set of input probabilities whereas for the latter the set of input probabilities changes with the receipt of every output symbol. The knowledge of every received y_j gives rise to constructing a D-ary code for the input alphabet; so if the output consists of s symbols we have to construct s D-ary codes and use the jth code to encode a symbol from the input if the received symbol is y_j. This tantamounts to choosing a sequence of code words from the known sequence of s codes. For this purpose all these s codes should be instantaneous because if they are taken to be uniquely decipherable codes then it is not known that the above choice of code words from these s uniquely decipherable codes will result in a uniquely decipherable code. So that generalisation is true only for instantaneous codes.

In this section we aim at proving a same sort of generalisation taking the Renyi entropy given by (33) as the basic measure of uncertainty. As pointed out in the section 1 that our generalisation is a generalisation of Campbell coding theorem and since Campbell coding theorem is a generalisation

of Shannon's first fundamental theorem so our generalisation is also a generalisation of Shannon's first fundamental theorem.

For this we need the concept of α-equivocation, $\alpha > 0$, $\alpha \neq 1$, associated with a channel when the channel input entropy is measured by (33).

By (6) of chapter 2 this is given by

$$H_{\alpha}(X/Y) = (1/1-\alpha) \sum_{j} p(y_j) \log \sum_{i} p^{\alpha}(x_i /y_j) \tag{34}$$

We also know from the proof of lemma of section 3.2 that for an nth extension of a channel with input probability distribution as a product distribution

$$H_{\alpha}(X^n /Y^n) = n H_{\alpha}(X/Y) \tag{35}$$

Campbell (1965) also showed that

$$(1/t)\log_D \sum_{i} p(x_i) D^{tn_i} \geq (1/1-\alpha) \log \sum_{i} p^{\alpha}(x_i) \tag{36}$$

where $(1/1+t) = \alpha$.

As in the generalisation of Shannon's first theorem referred to in the above paragraph and on the lines of which our generalisation for Campbell coding theorem is based, we shall consider s D-ary codes corresponding to s outputs. We shall assume that these s codes are instantaneous and their word lengths are given by

Table 1

Input symbols	Code 1	Code 2 ----------	Code s
x_1	n_{11}	n_{12} --------	n_{1s}
x_2	n_{21}	n_{22} --------	n_{2s}
---	---	---	---
x_r	n_{r1}	n_{r2} --------	n_{rs}

The probability distribution attached to the jth code is $\left[p(x_1/y_j), \ldots\ldots, p(x_r/y_j)\right]$.

So for the jth code, (36) reduces to

$$(1/t)\log \sum_i p(x_i/y_j)\, D^{tn_{ij}} \geqslant (1/1-\alpha)\log \sum_i p^{\alpha}(x_i/y_j)$$

i.e. $$L^{j}(t) \geqslant H_{\alpha}(X/y_j) \qquad (37)$$

Denoting the cost of transmitting an input symbol by $L^{(1)}(t)$, we have from (37) ,

$$L^{(1)}(t) = \sum_j p(y_j)\, L^{j}(t) \geqslant H_{\alpha}(X/Y) \qquad 0 < t < \infty \qquad (38)$$

As laid down in table 1 above, n_{ij} is the length of code word for the input symbol x_i when y_j is the output, so choosing n_{ij} so that

$$- \alpha \log p(x_i/y_j) + \log \sum_i p^{\alpha}(x_i/y_j) \leq n_{ij} \tag{39}$$
$$< - \alpha \log p(x_i/y_j) + \log \sum_i p^{\alpha}(x_i/y_j) + 1$$

where base of the logs is taken to be D. (39) can be written as

$$\log p^{-\alpha}(x_i/y_j) \sum_i p^{\alpha}(x_i/y_j) \leq \log D^{n_{ij}} < \log D\, p^{-\alpha}(x_i/y_j) \sum_i p^{\alpha}(x_i/y_j)$$

or

$$p^{-\alpha t}(x_i/y_j)\left[\sum_i p^{\alpha}(x_i/y_j)\right]^t \leq D^{t\, n_{ij}} < D^t\, p^{-\alpha t}(x_i/y_j)\left[\sum_i p^{\alpha}(x_i/y_j)\right]^t \tag{40}$$

Multiplying (40) by $p(x_i/y_j)$, summing over the subscript i and using the fact that $(1/1+t) = \alpha$ we have

$$\left(\sum_i p^{\alpha}(x_i/y_j)\right)^{1+t} \leq \sum_i p(x_i/y_j) D^{tn_{ij}} < D^t\left(\sum_i p^{\alpha}(x_i/y_j)\right)^{1+t} \tag{41}$$

In (41), first taking logarithms to the base D and then dividing by t we have,

$$H_{\alpha}(X/y_j) \leq L^{j}(t) < H_{\alpha}(X/y_j) + 1 \tag{42}$$

Multiplying (42) by $p(y_j)$ and summing over j we have

$$H_{\alpha}(X/Y) \leq L^{(1)}(t) < H_{\alpha}(X/Y) + 1$$

Similarly for an nth extension defined in section 3.2 we have

$$H_{\alpha}(X^n/Y^n) \leq L_n^{(1)}(t) < H_{\alpha}(X^n/Y^n) + 1 \tag{43}$$

Using (35), (43) reduces to

$$H_\alpha(X/Y) \leq \frac{L_n^{(1)}(t)}{n} < H_\alpha(X/Y) + \frac{1}{n} \tag{44}$$

The attainment of bound in (38) follows from taking limit of (44) as $n \to \infty$.

If $t = 0$ this is just the generalization given in Abramson (1963). If $t = \infty$ we choose each n_{ij} for all i to satisfy

$$\log r \leq n_{ij} < \log r + 1 \tag{45}$$

Since $\mathop{Lt}_{t\to\infty} L^j(t) = L^j(\infty) = \max_i n_{ij}$, so (45) can be written as

$$\log r \leq L^j(\infty) < \log r + 1 \tag{46}$$

Multiplying (46) by $p(y_j)$ and summing over j we have

$$H_0(X/Y) \leq L^{(1)}(\infty) < H_0(X/Y) + 1$$

where $H_0(X/Y)$ is the value of $H_\alpha(X/Y)$ at $\alpha = 0$.

Similarly for an nth extension we shall have

$$H_0(X^n/Y^n) \leq L_n^{(1)}(\infty) < H_0(X^n/Y^n) + 1 \tag{47}$$

Using (35), (47) reduces to

$$H_0(X/Y) \leq \frac{L_n^{(1)}(\infty)}{n} < H_0(X/Y) + 1/n \tag{48}$$

the result follows from (48) in the limit as $n \to \infty$.

This completes the description of a coding procedure for attaining the bound in (38).

5. α- BOUNDS , SOME RESULTS AND PROBLEMS

5.1 α-Bounds

Let

$$\bar{L} = \sum_i p(x_i)\, n_i \tag{49}$$

where $\sum_i p(x_i) = 1$ and $p(x_i)$ is the probability of transmitting through a noiseless channel from a uniquely decipherable code a code word of length n_i for the ith input symbol. Shannon gave the following lower bound for $\bar{L}$

$$\bar{L} \geqslant \frac{H_1(X)}{\log D}$$

where $H_1(X)$ is the Shannon's measure of entropy and D is the no. of letters in the encoding alphabet. The following theorem generalises this bound.

Th. 1 Let $X = [x_1, x_2, \ldots, x_r]$ be a message ensemble from a memoryless source with respective probabilities of transmission given by $[p(x_1), \ldots, p(x_r)]$ so that $\sum_{i=1}^{r} p(x_i) = 1$. If X is encoded in a sequence of uniquely decipherable words of lengths $[n_1, n_2, \ldots, n_r]$ from an encoding alphabet of D letters then

$$\text{(i)} \qquad \bar{L} = \sum_{i=1}^{r} p(x_i)\, n_i \geqslant \frac{H_\alpha(X)}{\log D} \qquad \alpha \geqslant 1 \tag{50}$$

and

(ii) If in particular $p(x_i) = D^{-n_i}/\sum_i D^{-n_i}$ then for $0 < \alpha < 1$

$$\bar{L} \leq \frac{H_\alpha(X)}{\log D} - (\log \sum_i D^{-n_i})/\log D \tag{51}$$

__Proof.__ (i)

By Jensen's inequality

$$\log \sum_i p^\alpha(x_i) = \log \sum_i p(x_i) p^{\alpha-1}(x_i) \geq \sum_i p(x_i) \log p^{\alpha-1}(x_i) \tag{52}$$

We know that if $[p(x_i)]$ and $[q_i]$ are two sets of non-negative numbers such that $\sum_i p(x_i) = \sum_i q_i = 1$ then

$$-\sum_i p(x_i) \log p(x_i) \leq -\sum_i p(x_i) \log q_i \tag{53}$$

For $\alpha > 1$, (53) can be written as

$$\sum_i p(x_i) \log p^{\alpha-1}(x_i) \geq \sum_i p(x_i) \log q_i^{\alpha-1} \tag{54}$$

Putting $q_i = D^{-n_i}/\sum_i D^{-n_i}$ in (54) and using (52) we shall have

$$\log \sum_i p^\alpha(x_i) \geq \bar{L}(1-\alpha) \log D - (\alpha-1) \log \sum_i D^{-n_i} \tag{55}$$

Dividing (55) by $(1/1-\alpha)$, a negative no., and using the result of McMillan's theorem that $\sum_i D^{-n_i} \leq 1$ for a uniquely decipherable code we shall have

$$\bar{L} \geq H_\alpha(X) / \log D \qquad \alpha > 1, \tag{56}$$

Since this result is already known to be true for $\alpha = 1$, so combining this with (56) we shall have (50).

It may be noted that this result also follows from the known result and the monotonic decreasing character of $H_\alpha(X)$, $\alpha > 0$, proved in section 2.2 of chapter 2. The above is an independent proof.

(ii)

Since $p(x_i) = \dfrac{D^{-n_i}}{\sum_i D^{-n_i}}$, so

$$\log \sum_i p^\alpha(x_i) = \log \sum_i p(x_i) \left(\frac{D^{-n_i}}{\sum_i D^{-n_i}} \right)^{\alpha-1}$$

$$= \log \sum_i p(x_i) D^{-n_i(\alpha-1)} - (\alpha-1) \log \sum_i D^{-n_i} \qquad (57)$$

But by Jensen's inequality

$$\log \sum_i p(x_i) D^{-n_i(\alpha-1)} \geqslant \sum_i p(x_i) \log D^{-n_i(\alpha-1)}$$

$$= -(\alpha-1)\, \bar{L} \log D \qquad (58)$$

Using (58), (57) reduces to

$$\log \sum_i p^\alpha(x_i) \geqslant -(\alpha-1)\, \bar{L} \log D - (\alpha-1) \log \sum_i D^{-n_i} \qquad (59)$$

Multiplying (59) by $(1/1-\alpha)$, a positive number, we shall have (51).

Q.E. D.

5.2 Results that Follow

(i) It is mentioned on page 140, Reza (1961) that ''If the requirement of successive equiprobable partitioning of the probability matrix of a message ensemble is approximately satisfied then reasonably efficient encoding procedures can be expected.''

The proof of this statement, when $p(x_i) = D^{-n_i}/\sum_i D^{-n_i}$, follows from the above theorem.

Proof. When

$$p(x_i) = D^{-n_i} / \sum_i D^{-n_i} \tag{60}$$

from (50) and (51) we have in the limit as $\alpha \to 1$

$$\frac{H_1(X)}{1-} \leq \bar{L} \leq \frac{H_1(X)}{\log D} - \frac{\log \sum_i D^{-n_i}}{\log D} \tag{61}$$

that if n_i subject to (60) are chosen $\sum_i D^{-n_i}$ becomes almost or actually
:istence of a uniquely decipherable
which the average word length $\bar{L}$ would
equal to $H_1(X)/\log D$. The nearer
$_i$ = 1 we are in respect of our choice
) the better would be the attainment
:ing the average word length achieve
/ log D and better the approximation
the requirement of successive equi-
f the probability matrix being met
for an efficient code.

Please do not displace.

Thank you

Q. E. D.

(ii) ''What set of source probabilities for a uniquely decipherable code would achieve the lower bound $H_1(X)/\log D$ for $\bar{L}$ in noiseless coding ?'' was one time a baffling question. Now its complete solution is given by the theorem

'' The necessary and sufficient condition for $\bar{L}$ to be equal to $H_1(X) / \log D$ is that $p(x_i) = D^{-n_i}$. ''

The sufficient condition of this theorem also follows from the above theorem when $p(x_i) = D^{-n_i}/\sum_i D^{-n_i}$.

Proof. When $p(x_i) = D^{-n_i}/\Sigma D^{-n_i}$, both the results (50) and (51) of the above theorem are valid . (51) gives a series of upper bounds and in the limit as $\alpha \to 1$ we shall have the least upper bound for $\bar{L}$ and this is given by, denoting it by l.u.b. $\bar{L}$,

$$\text{l.u.b. } \bar{L} = \frac{H_1(X)}{\log D} - \frac{\log \Sigma D^{-n_i}}{\log D} \tag{62}$$

and for $\alpha = 1$ the known greatest lower bound for $\bar{L}$, denoting it by g.l.b. $\bar{L}$, is

$$\text{g.l.b. } \bar{L} = \frac{H_1(X)}{\log D} \tag{63}$$

From (62) and (63) it immediately follows that

$$\text{l.u.b. } \bar{L} = \text{g.l.b. } \bar{L} \quad \text{iff } \sum_i D^{-n_i} = 1 \tag{64}$$

Hence from $p(x_i) = D^{-n_i} / \sum_i D^{-n_i}$ and (64) it follows that

$$p(x_i) = D^{-n_i} \Rightarrow \bar{L} = \frac{H_1(X)}{\log D}$$

Q.E.D.

5.3 Problems that Arise

(i) Campbell (1965) defined a cost function of order t

$$L(t) = (1/t)\ \frac{1}{t} \log \sum_i p(x_i)\, D^{tn_i} \qquad 0< t< \infty, \qquad (65)$$

where $\frac{1}{1+t} = \alpha$ and D is in the number of letters in an encoding alphabet used for constructing a uniquely decipherable code and n_i is the length of a code word for the source symbol x_i to be transmitted through a noiseless channel with probability $p(x_i)$ and proved a coding theorem for noiseless channels by taking Renyi's entropy given by (33) as the basic measure of uncertainty. In the limit, as $\alpha \to 1$, Campbell's result tends to the corresponding Shannon's result.

For $0< \alpha < 1$, multiplying (52) by $(1/1-\alpha)$, a positive number, we shall have

$$H_\alpha(X) \geqslant H_1(X) \qquad 0< \alpha < 1 \qquad (66)$$

Equality in (66) is possible when all the x_i are equiprobable. When the cost of transmitting a code word is measured by (50), the Shannon's result says that by taking sufficiently large nth extension of the source it is possible to encode the symbols of the so extended source into a uniquely decipherable code such that the difference between $\bar{L}_n /n$ and $H_1(X) /\log D$ is as small as desired.

Campbell's result is restricted to $0< \alpha < 1$ since the cost function (65) is true for $0< t < \infty$ i.e. for $0< \alpha < 1$

(by $\alpha = \frac{1}{1+t}$). By virtue of (66) and the above Shannon's result it seems feasible to have a uniquely decipherable code for an nth extension of the source so that the difference between $\bar{L}_n/n$ and $H_\alpha(X) / \log D$ is almost negligible when the entropy of the source is measured by the corresponding member of $\{H_\alpha(X),\ 0 < \alpha < 1\}$.

The beauty of cost function (65) lies in the fact that Campbell by considering it was not only able to give a lower bound for $L(t)$ in terms of $H_\alpha(X)$ but also gave a method of choosing n_i for the purpose. But a method, as to how to choose n_i so that the resultant code is a uniquely decipherable code for an nth extension and the difference between $\bar{L}_n/n$ and $H_\alpha(X) / \log D$ is as small as desired, is still to be developed.

(ii) Since $H_\alpha(X) \leq H_1(X)$ for $\alpha > 1$, the result in (4) gives us a better lower bound of $\bar{L}$ for uniquely decipherable codes if we measure the entropy of the source by a member of $\{H_\alpha(X),\ \alpha > 1\}$.

For the already available lower bound

$$\bar{L} \geq \frac{H_1(X)}{\log D}$$

We know that by choosing n_i so that

$$\frac{-\log p(x_i)}{\log D} \leq n_i < \frac{-\log p(x_i)}{\log D} + 1$$

we have not only the guarantee of the existence of a uniquely

decipherable code but can also make $\bar{L}$ achieve its lower bound.

In this case not only a method of as to how to choose n_i in practice with a view to have a uniquely decipherable code so that the bound in (50) for $\alpha > 1$ is more or less or actually achieved but also the reasoning for visualising its very feasibility are to be developed.

It may be noted that all the results developed in this chapter are true also for the alternative definitions of $H_\alpha(Y/X)$ and $H_\alpha(X/Y)$ given respectively by (100) and (101) of chapter 2.

CHAPTER -5

α - ENTROPIES FOR CONTINUOUS MEMORYLESS CHANNELS

This chapter consists of two parts. Part I deals with bivariate continuous memoryless channels and part II with multivariate continuous memoryless channels.

PART - I

1. INTRODUCTION AND EXTENSION OF RENYI'S α-ENTROPY

As given in chapter 1 Renyi's α-entropy for a complete discrete scheme with probability matrix

$$\left[p(x_i)\right] \; i=1,2,\dots,n \; , \; p(x_i) \geqslant 0 \; , \; \sum_i p(x_i) = 1$$

is given by

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2 \sum_i p^\alpha(x_i) \tag{1}$$

(1) can be written as

$$-\log_2 \left[\sum_i p(x_i)\, p^{\alpha-1}(x_i)\right]^{\frac{1}{\alpha-1}}$$

or

$$-\log_2 M_{\alpha-1}$$

where

$$M_{\alpha-1} = \left[\sum_i p(x_i)\, p^{\alpha-1}(x_i)\right]^{\frac{1}{\alpha-1}} \tag{2}$$

is known to be a power mean of order $\alpha-1$ of n positive numbers $p(x_1), p(x_2), \ldots\ldots, p(x_n)$.

Corresponding to (2) the power mean of order $\alpha-1$ for a complete continuous probability scheme having the probability density $f(x)$ is

$$\left[\int_{-\infty}^{\infty} f(x)\ f^{\alpha-1}(x)\ dx\right]^{\frac{1}{\alpha-1}} \tag{3}$$

where $\int_{-\infty}^{\infty} f(x)\ dx = 1$. So following the pattern of the structure of (1) we may define entropy of order α, $\alpha > 0$, $\alpha \neq 1$ for such a scheme as

$$H_\alpha(X) = -\log_2 \left[\int_{-\infty}^{\infty} f(x)\ f^{\alpha-1}(x)\ dx\right]^{\frac{1}{\alpha-1}}$$

$$= \frac{1}{1-\alpha} \log_2 \int_{-\infty}^{\infty} f^{\alpha}(x)\ dx \tag{4}$$

Another way of arriving at (4) is as follows:

Let a continuous scheme have $f(x)$ as its probability density so that $\int_{-\infty}^{\infty} f(x)\ dx = 1$. If $f(x)$ is finite and positive everywhere in its domain of definition then the Kolmogorov- Nagumo generalised mean of $\log_2 \frac{1}{f(x)}$ by taking $x = \log_2 \frac{1}{f(x)}$ in the strictly monotonic and continuous parametric function

$$g_{\alpha}(x) = 2^{(1-\alpha)x} \qquad \alpha > 0, \alpha \neq 1 \qquad (5)$$

is

$$g_{\alpha}^{-1}\left[\int_{-\infty}^{\infty} f(x) \quad g_{\alpha}(\log_2 \frac{1}{f(x)}) \, dx\right]$$

$$= \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} f^{\alpha}(x) \, dx \qquad (6)$$

It may be noted that if we adopt Shannon's convention that $0 \times \infty = 0$ or evaluate (6) in the extended real number system, then (6) is also defined even when $f(x)$ vanishes at the most over a set of measure zero.

The power mean of order $\alpha-1$ given by (3) has extensively been studied. The following are a few of its properties that we need for our subsequent development:

(i) it is a continuous functions of α, page 13, Littlewood (1944).

(ii) it is a monotonic increasing function of α, page 144, Hardy, Littlewood and Polya (1952).

By property (i) for (3), $H_{\alpha}(X)$ defined by (4) is also a continuous function of α and by property (ii) for (3) it is a monotonic decreasing function of α.

It can be easily verified that in the limit, as $\alpha \to 1$, (4) tends to the corresponding Shannon's definition.

Taking (4) as the basic measure of α-uncertainty, other α-entropies associated with a two-dimensional continuous memoryless channel are defined, some relations involving these α-entropies are developed ~~some relations involving~~

~~these α-entropies are developed~~ and the drawbacks of $H_\alpha(X)$ used as a measure of entropy for a continuous scheme are discussed. All the results developed below tend in the limit, as $\alpha \to 1$, to the corresponding available results.

The main tools used are Jensen's and Holder's inequalities for integrals. The case, when the underlying schemes are independent, has been included while applying Hölder's inequality and in this case there will be strict equality. In the case of Jensen's inequality, the equality will hold when the density function, of whose convex or concave function the inequality deals with, is uniformly distributed.

All the operations, dealt with below, are contingent upon the existence of underlying integrals. From section 3 onwards, base of the logs will be assumed to be e.

2. α- Entropies Associated With a Two Dimensional Continuous Channel.

Suppose a two-dimensional continuous random variable has $f(x,y)$ - the joint probability density -, $f(x)$ and $f(y)$ the marginal probability densities, $f(y/x)$ the conditional probability density of Y given a value x of X- and $f(x/y)$ the conditional prob. density of X given a value y of Y so that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1, \int_{-\infty}^{\infty} f(x,y)\, dy = f(x), \int_{-\infty}^{\infty} f(x,y)dx = f(y)$$

$$f(y/x) = \frac{f(x,y)}{f(x)} \quad , \quad f(x/y) = \frac{f(x,y)}{f(y)}$$

There are five α-entropies, $\alpha > 0$, $\alpha \neq 1$, associated with such a variable.

(i) $H_\alpha(X)$, defined for the input X. It is given by (4).

(ii) $H_\alpha(Y)$, defined for the output Y. It can similarly be defined as $H_\alpha(X)$.

(iii) $H_\alpha(X,Y)$, defined for the system as a whole, Based on the procedure adopted for getting (4), it is given by

$$H_\alpha(X,Y) = \frac{1}{1-\alpha} \log_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^\alpha(x,y)\, dx\, dy \qquad (7)$$

(iv) $H_\alpha(Y/X)$, defined about the output Y when it is known that X is transmitted and may be termed as conditional α-entropy. It may be defined as such:

Based on the procedure for getting (4) conditional α-entropy calculated on the assumption that the event x of X has occurred is

$$H_\alpha(Y/x) = \frac{1}{1-\alpha} \log_2 \int_{-\infty}^{\infty} f^\alpha(y/x)\, dy \qquad \alpha > 0,\ \alpha \neq 1 \qquad (8)$$

To have a measure for $H_\alpha(Y/X)$ it is quite natural to take the expected value of (8); so

$$H_\alpha(Y/X) = \frac{1}{1-\alpha} \int_{-\infty}^{\infty} f(x) \left[\log_2 \int_{-\infty}^{\infty} f^\alpha(y/x)\, dy\right] dx, \quad \alpha > 0, \alpha \neq 1 \quad (9)$$

(v) $H_\alpha(X/Y)$, defined about the input X when it is known that Y is received and may be termed as α- equivocation. Based on the procedure adopted for getting (9), it is given by

$$H_\alpha(X/Y) = \frac{1}{1-\alpha} \int_{-\infty}^{\infty} f(y) \left[\log_2 \int_{-\infty}^{\infty} f^\alpha(x/y)\, dx\right] dy \quad \alpha > 0, \alpha \neq 1 \quad \ldots (10)$$

In the limit, as $\alpha \to 1$, (7), (9) and (10) tend to the corresponding Shannon's definitions.

Since

$$\underset{\alpha \to \infty}{\mathrm{Lt}} \left[\int_{-\infty}^{\infty} f(x)\, f^{\alpha-1}(x)\, dx\right]^{\frac{1}{\alpha-1}} = \max.\ f(x)$$

[Page 143 Hardy, Littlewood and Polya (1952)]

so (i) $\underset{\alpha \to \infty}{\mathrm{Lt}}\ H_\alpha(X) = -\log_2 \max.\ f(x)$

(ii) $\underset{\alpha \to \infty}{\mathrm{Lt}}\ H_\alpha(X,Y) = -\log_2 \max.\ f(x,y)$

(iii) $\underset{\alpha \to \infty}{\mathrm{Lt}}\ H_\alpha(Y/X) = -\int_{-\infty}^{\infty} f(x) \left[\log_2 \max_y . f(y/x)\right] dx$

3. Some Relations Between α-Entropies

3.1 Relation between $H_\alpha(Y/X)$ and $H_\alpha(Y)$

Case when $0 < \alpha < 1$

Since log x is concave, so taking $x = \int_{-\infty}^{\infty} f^{\alpha}(y/x)dy$ in

$$\int_{-\infty}^{\infty} f(x)\left[\log \int_{-\infty}^{\infty} f^{\alpha}(y/x)\, dy\right] dx$$

and applying Jensen's inequality for integrals Page 151, Hardy, Littlewood and Polya (1952), we have

$$\int_{-\infty}^{\infty} f(x)\left[\log \int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right] dx \leq \log \int_{-\infty}^{\infty} f(x)\left[\int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right] dx$$

$$= \log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{1-\alpha}(x)\, f^{\alpha}(x) f^{\alpha}(y/x)dx\, dy$$

$$= \log \int_{-\infty}^{\infty} f^{\alpha}(y)\left[\int_{-\infty}^{\infty} f^{\alpha}(x/y) f^{1-\alpha}(x)dx\right] dy \qquad \ldots (11)$$

But in

$$\int_{-\infty}^{\infty} f^{\alpha}(x/y)\, f^{1-\alpha}(x)\, dx$$

taking $k = (1/\alpha) > 1$ and $k' = (1/1-\alpha)$ so that $(1/k)+(1/k') = 1$ and applying Holder's inequality for integrals for these value of the parameter k, page 140 Hardy, Littlewood and Polya (1952), we have

$$\int_{-\infty}^{\infty} f^{\alpha}(x/y)\, f^{1-\alpha}(x)\, dx \leq \left[\int_{-\infty}^{\infty} f(x/y)dx\right]^{\alpha} \left[\int_{-\infty}^{\infty} f(x)\, dx\right]^{1-\alpha} = 1$$

Let $\max\limits_{y} \int_{-\infty}^{\infty} f^{\alpha}(x/y)\, f^{1-\alpha}(x)\, dx = \beta \leq 1$. By this (11)

reduces to

$$\int_{-\infty}^{\infty} f(x) \left[\log \int_{-\infty}^{\infty} f^{\alpha}(y/x)\, dy\right] dx \leq \log \beta + \log \int_{-\infty}^{\infty} f^{\alpha}(y)dy$$

Multiplying this by $(1/1-\alpha)$, a positive no., and using the fact that $\beta \leq 1$ we have

$$H_{\alpha}(Y/X) \leq H_{\alpha}(Y) \qquad 0 < \alpha < 1 \qquad (12)$$

<u>Case when $\alpha > 0$</u>

Since $$\int_{-\infty}^{\infty} f^{\alpha}(y/x)\, dy \leq \left[\int_{-\infty}^{\infty} f\,(y/x)\, dy\right]^{\alpha} = 1, \qquad (13)$$

so

$$\int_{-\infty}^{\infty} f(x) \left[\log \int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right] dx \geq \int_{-\infty}^{\infty} f(x) \left[\left\{\int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right\} \log \left\{\int_{-\infty}^{\infty} f(y/x)dy\right\}\right] dx \quad \dots. (14)$$

Since $x \log x$ is convex, so taking $x = \int_{-\infty}^{\infty} f^{\alpha}(y/x)dy$ on the R.H.S. of (14) and applying Jensen's inequality for integrals, we have

$$\text{R.H.S. of (14)} \geq \left[\int_{-\infty}^{\infty} f(x) \left\{\int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right\} dx\right] \log \left[\int_{-\infty}^{\infty} f\,(x) \left\{\int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right\} dx\right] \quad \dots \quad (15)$$

But by (13) $\int_{-\infty}^{\infty} f^{\alpha}(y/x)\, dy \leq 1$, so taking

$\min_{x} \int_{-\infty}^{\infty} f^{\alpha}(y/x)\, dy = \beta \leq 1$, (15) reduces to

$$\text{R.H.S.}(14) \geqslant \beta \log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\, f^{\alpha}(y/x)\, dx\, dy$$

$$= \beta \log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{1-\alpha}(x)\, f^{\alpha}(y)\, f^{\alpha}(x/y)\, dx\, dy$$

$$= \beta \log \int_{-\infty}^{\infty} f^{\alpha}(y) \left[\int_{-\infty}^{\infty} f^{\alpha}(x/y)\, f^{1-\alpha}(x)\, dx\right] dy \qquad (16)$$

But in

$$\int_{-\infty}^{\infty} f^{1-\alpha}(x)\, f^{\alpha}(x/y)\, dx$$

ing $k = 1/\alpha < 1$ and $k' = (1/1-\alpha)$ so that $(1/k)+(1/k')=1$ applying Hölder inequality for integrals for these values the parameter k, page 140, Hardy, Littlewood and Polya 52), we have

$$\int f^{1-\alpha}(x)\, f^{\alpha}(x/y)\, dx \geqslant \left[\int_{-\infty}^{\infty} f(x)dx\right]^{1-\alpha} \left[\int_{-\infty}^{\infty} f(x/y)dx\right]^{\alpha} = 1$$

$\int_{-\infty}^{\infty} f^{1-\alpha}(x)\, f^{\alpha}(x/y)dx = \gamma \geqslant 1.$ By this and (14),(16)

Multiplying (17) by $(1/1-\alpha)$, a negative no., and using the fact that $\beta \leqslant 1$ and $\gamma \geqslant 1$, we have

$$H_\alpha(Y/X) \leqslant H_\alpha(Y) \qquad \alpha > 1 \qquad (18)$$

Combining (12) and (18) we have

$$H_\alpha (Y/X) \leqslant H_\alpha(Y) \qquad \alpha > 0, \alpha \neq 1 \qquad (19)$$

In the limit, as $\alpha \to 1$, (19) tends to the Shannon's inequality $H_1(Y/X) \leqslant H_1(Y)$, so combining this symbolically with (19) we have

$$H_\alpha(Y/X) \leqslant H_\alpha(Y) \qquad \alpha > 0 \qquad (20)$$

Similarly $H_\alpha(X/Y) \leqslant H_\alpha(X) \qquad \alpha > 0 \qquad (21)$

3.2 Relations between $H_\alpha(X,Y)$, $H_\alpha(X)$ and $H_\alpha(Y/X)$

Two infinite sets of inequalities, every member of either set tending in the limit, as $\alpha \to 1$, to the known Shannon's iequality $H_\alpha(X,Y) = H_\alpha(X) + H_\alpha(Y/X)$ and two bounds for $H_\alpha(X,Y)$, one upper for $0 < \alpha < 1$ and one lower for $\alpha > 1$, are developed.

Case (i) for $0 < \alpha < 1$

$$\log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^\alpha(x,y)dx\, dy = \log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^\alpha(x)\, f^\alpha(y/x)\, dx\, dy$$

$$= \log \int_{-\infty}^{\infty} f(x)^{\alpha-\frac{1}{k}} \left[\int_{-\infty}^{\infty} f^{1/k}(x) f^\alpha(y/x)\, dy \right] dx \qquad (22)$$

where $0 < k < 1$.

For $0 < k < 1$ and k' chosen subject to $(1/k)+(1/k')=1$ we know by Holder's inequality

$$\int_{-\infty}^{\infty} \phi(x)\, \phi_2(x)dx \geqslant \left[\int_{-\infty}^{\infty} \phi_1^{k}(x)\, dx\right]^{1/k} \left[\int_{-\infty}^{\infty} \phi_2^{k'}(x)dx\right]^{1/k'} ;$$

Taking $\phi_1(x) = \int_{-\infty}^{\infty} f^{1/k}(x)\, f^{\alpha}(y/x)dy$ and $\phi_2(x) = f^{\alpha - 1/k}(x)$

in it we have

$$\int_{-\infty}^{\infty} f(x)^{\alpha-1/k}\left[\int_{-\infty}^{\infty} f(x)^{1/k}\, f^{\alpha}(y/x)\right] dx \geqslant \left[\int_{-\infty}^{\infty} f(x)^{(\alpha-1/k)k'} dx\right]^{1/k'} \tag{23}$$

$$\left[\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} f(x)^{1/k} f^{\alpha}(y/x)dy\right\}^{k} dx\right]^{1/k}$$

Using (23), (22) reduces to

$$\log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^{\alpha}(x,y)dx\, dy \geqslant \frac{1}{k'} \log \int_{-\infty}^{\infty} f^{(\alpha-1/k)k'}(x)\, dx + 1/k$$

$$\log \int_{-\infty}^{\infty} f(x)\left\{\int_{-\infty}^{\infty} f^{\alpha}(y/x)\, dy\right\}^{k} dx \tag{24}$$

But by Jensen's inequality

$$(1/k) \log \int_{-\infty}^{\infty} f(x)\left\{\int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right\}^{k} dx \geqslant$$

$$\int_{-\infty}^{\infty} f(x)\left\{\log \int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right\} dx$$

Using this in (24) and then dividing it thro'out by $(1/1-\alpha)$, a positive no., we have

$$H_\alpha(X,Y) \geqslant H_{(\alpha-1/k)k'}(X) + H_\alpha(Y/X) \tag{25}$$

where $0 < \alpha < 1$, $0 < k < 1$, and $1/k + 1/k' = 1$

But for $0 < \alpha < 1$, $0 < k < 1$

$$\frac{1-\alpha k}{1-k} > 1 ,$$

so $(\alpha - 1/k)k' = (1-\alpha k)/(1-k)$ lie between 1 and ∞. Hence (25) gives an infinite no. of inequalities connecting $H_\alpha(X,Y)$ and $H_\alpha(Y/X)$ for $0<\alpha<1$ to $H_\alpha(X)$ for $\alpha > 1$. It can be easily verified that in the limit, as $\alpha \to 1$, all these inequalities tend to the Shannon's equality $H_1(X,Y) = H_1(X) + H_1(Y/X)$.

<u>Case (ii) for $0<\alpha<1$</u>

Supposing $k > 1$ in (22), choosing k' subject to $1/k + 1/k' = 1$ and applying Holder inequality for this range of the parameter k we have

$$\log \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^\alpha(x,y)\,dxdy \leqslant \log \left[\int_{-\infty}^{\infty} \left\{ f(x)^{\alpha-1/k} \right\}^{k'} dx \right]^{1/k'}$$

$$+ \log \left[\int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f^\alpha(y/x)\, f(x)^{1/k}\, dy \right\}^{k} dx \right]^{1/k} \tag{26}$$

Multiplying (26) by $(1/1-\alpha)$, a positive number, we have

$$H_\alpha(X,Y) \leqslant H_{(\alpha-1/k)k'}(X) + \frac{1}{(1-\alpha)k} \log \int_{-\infty}^{\infty} f(x) \left\{ \int_{-\infty}^{\infty} f^\alpha(y/x)\,dy \right\}^{k} dx \tag{27}$$

where $0<\alpha<1$, $k > 1$ and $1/k + 1/k' = 1$

Taking limits of (27) as $k \to \infty$ we have

$$\underset{k \to \infty}{\mathrm{Lt}} \; H_{(\alpha-1/k)k'}(X) = H_\alpha(X) \tag{28}$$

$$\underset{k \to \infty}{\mathrm{Lt}} \frac{1}{(1-\alpha)k} \log \int_{-\infty}^{\infty} f(x) \left\{ \int_{-\infty}^{\infty} f^\alpha(y/x)\, dy \right\}^k dx$$

$$= \underset{k \to \infty}{\mathrm{Lt}} \frac{\int_{-\infty}^{\infty} f(x) \left\{ \int_{-\infty}^{\infty} f^\alpha(y/x) dy \right\}^k \left\{ \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} f^\alpha(y/x) dy \right\} dx}{\int_{-\infty}^{\infty} f(x) \left\{ \int_{-\infty}^{\infty} f^\alpha(y/x)\, dy \right\}^k dx} \tag{29}$$

Since the R.H.S. of (29) is a weighted mean of

$\frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} f^\alpha(y/x)\, dy = H_\alpha(Y/x)$, so L.H.S. of (29)

$\leq \max_x H_\alpha(Y/x)$. Using this and (28), (27) reduces to

$$H_\alpha(X,Y) \leq H_\alpha(X) + \max_x H_\alpha(Y/x) \tag{30}$$

(30) gives an upper bound for $H_\alpha(X,Y)$ for $0 < \alpha < 1$.

<u>Case (i) for $\alpha > 1$</u> .

In deriving (24) we did not make any condition on α, so applying Jensen's inequality to the 2nd term of its R.H.S. and then multiplying it thro'out by $(1/1-\alpha)$, a negative no., we have

$$H_\alpha(X,Y) \leq H_{(\alpha-1/k)k'}(X) + H_\alpha(Y/X) \tag{31}$$

where $\alpha > 1$, $0 < k < 1$ and $1/k + 1/k' = 1$.

Under the conditions for which (31) is valid, $(\alpha-1/k)k' < 1$, so (31) gives an infinite no. of inequalities connecting $H_\alpha(X,Y)$ and $H_\alpha(Y/X)$ for $\alpha > 1$, $\cancel{\alpha \neq \infty}$ to $H_\alpha(X)$ for $0 < \alpha < 1$. Every member of this set tends, in the limit as $\alpha \to 1$, to the Shannon's equality $H_1(X,Y) = H_1(X) + H_1(Y/X)$.

<u>Case (ii) for $\alpha > 1$</u>

Again in deriving (26) no condition on α was made, so multiplying it thro'out by $(1/1-\alpha)$, a negative number, we have

$$H_\alpha(X,Y) \geqslant H_{(\alpha-1/k)k'}(X) + \frac{1}{(1-\alpha)k} \log \int_{-\infty}^{\infty} f(x) \left\{ \int_{-\infty}^{\infty} f^\alpha(y/x)dy \right\}^k dx \qquad \text{.... (32)}$$

In the limit, as $k \to \infty$, (32) tends to

$$H_\alpha(X,Y) \geqslant H_\alpha(X) + \min_{x} H_\alpha(Y/x) \qquad (33)$$

(33) gives a lower bound for $H_\alpha(X,Y)$ for $\alpha > 1$.

Similarly we can prove, for $0 < \alpha < 1$,

$$H_\alpha(X,Y) \geqslant H_{(\alpha-1/k)k'}(Y) + H_\alpha(X/Y)$$

$$H_\alpha(X,Y) \leqslant H_\alpha(Y) + \max_{y} H_\alpha(X/y)$$

and, for $\alpha > 1$,

$$H_\alpha(X,Y) \leqslant H_{(\alpha-1/k)k'}(Y) + H_\alpha(X/Y)$$

$$H_\alpha(X,Y) \geqslant H_\alpha(Y) + \min_{y} H_\alpha(X/y) .$$

<u>A Study of Equality Between</u> $H_\alpha(X,Y)$ <u>and</u> $H_\alpha(X)+H_\alpha(Y/X)$

(i) If X and Y are jointly Guassian, the relation

$$X,Y) = H_\alpha(X) + H_\alpha(Y/X) = H_\alpha(Y) + H_\alpha(X/Y) \qquad \alpha > 0 \quad (34)$$

s. The proof of this follows from the corresponding lt for multidimensional continuous schemes dealt with ection 2.4, part II of this chapter.

(ii) If $f(y/x)$ is uniformly distributed for each x, iall have

$$H_\alpha(X,Y) = H_\alpha(X) + H_\alpha(Y/X) \qquad \alpha > 0 \qquad (35)$$

, for $\alpha > 0$, $\alpha \neq 1$,

$$H_\alpha(X,Y) = (1/1-\alpha) \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^\alpha(x,y)\, dx\, dy$$

$$= (1/1-\alpha) \log \left[\int_{-\infty}^{\infty} f^\alpha(x)\, dx \int_{-\infty}^{\infty} f^\alpha(y/x)\, dy \right]$$

$$= (1/1-\alpha) \log \int_{-\infty}^{\infty} f^\alpha(x) dx + (1/1-\alpha) \log \int_{-\infty}^{\infty} f^\alpha(y/x) dy$$

$$= \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} f^\alpha(x) dx + \frac{1}{1-\alpha} \int_{-\infty}^{\infty} f(x) (\log \int_{-\infty}^{\infty} f^\alpha(y/x) dy) dx$$

$$= H_\alpha(X) + H_\alpha(Y/X) \qquad \alpha > 0,\ \alpha \neq 1 \qquad (36)$$

) $= H_1(X) + H_1(Y/X)$ is known to be true, so ibolically with (36), we shall have (35).

Similarly if $f(x/y)$ is uniformly distributed for each y, we shall have

$$H_\alpha(X,Y) = H_\alpha(Y) + H_\alpha(X/Y) \qquad \alpha > 0 \tag{37}$$

(iii) In general

$$H_\alpha(X,Y) \neq H_\alpha(X) + H_\alpha(Y/X) \neq H_\alpha(Y) + H_\alpha(X/Y) \quad \alpha > 0, \alpha \neq 1 \tag{38}$$

Consider

$$f(x,y) = x\, e^{-x(y+1)} \qquad x > 0,\ y \geqslant 0$$

$$= 0 \qquad \text{elsewhere}$$

The marginal densities are

$$f(x) = e^{-x} \qquad x \geqslant 0$$

$$= 0 \qquad \text{elsewhere}$$

$$f(y) = \frac{1}{(1+y)^2} \qquad y \geqslant 0$$

$$= 0 \qquad \text{elsewhere}$$

and the conditional densities are

$$f(y/x) = x\ e^{-xy} \qquad \text{for all } x > 0$$

$$f(x/y) = (1+y)^2\ x\ e^{-x(y+1)} \quad \text{for all } y \geqslant 0$$

So

$$H_\alpha(X,Y) = \frac{1}{1-\alpha} \log \int_0^\infty \int_0^\infty x^\alpha e^{-x\,\alpha(y+1)}\, dx\, dy$$

$$= \frac{\alpha+2}{1-\alpha} \log \frac{1}{\alpha} + \frac{1}{1-\alpha} \log \Gamma(\alpha+1) \quad \alpha > 0, \alpha \neq 1 \qquad (39)$$

$$H_\alpha(X) = (1/1-\alpha) \log \int_0^\infty e^{-\alpha x}\, dx$$

$$= (1/1-\alpha) \log 1/\alpha \quad , \quad \alpha > 0, \alpha \neq 1 \qquad (40)$$

$$H_\alpha(Y) = (1/1-\alpha) \log \int_0^\infty \frac{dy}{(1+y)^{2\alpha}} \qquad (41)$$

The integrand in (41) is convergent if $2\alpha > 1$ i.e. $\alpha > 1/2$; so for $\alpha > 1/2, \alpha \neq 1$.

$$H_\alpha(Y) = (1/1-\alpha) \log (1/2\alpha-1) \quad , \quad \alpha > 1/2 , \alpha \neq 1 \qquad (42)$$

$$H_\alpha(Y/X) = (1/1-\alpha) \int_0^\infty e^{-x} \left[\log \int_0^\infty x^\alpha e^{-\alpha x y}\, dy \right] dx$$

It does not exist

$$\cancel{= \frac{1}{1-\alpha} \log \frac{1}{\alpha} - 1 \quad , \quad \alpha > 1, \alpha \neq 1} \qquad (43)$$

$$H(X/Y) = \frac{1}{1-\alpha} \int_0^\infty \frac{1}{(1+y)^2} \left[\log \int_0^\infty x^\alpha e^{-\alpha x(y+1)} (1+y)^{2\alpha}\, dx \right] dy$$

$$= \frac{1}{1-\alpha} \int_0^\infty \frac{1}{(1+y)^2} \log \left[\Gamma(\alpha+1)\, \alpha^{-\alpha-1} (1+y)^{\alpha-1} \right] dy$$

$$= \frac{1}{1-\alpha} \log \Gamma(\alpha+1) + \frac{\alpha+1}{\alpha-1} \log \alpha - 1 \quad , \quad \alpha > 0, \alpha \neq 1$$

$$\dots \quad (44)$$

From (39), (40) and (43) we have in general for $\alpha > 0$, $\alpha \neq 1$

$$H_\alpha(X,Y) \neq H_\alpha(X) + H_\alpha(Y/X) \qquad (45)$$

and from (39), (42) and (44) we have in general for $\alpha > 1/2$, $\alpha \neq 1$

$$H_\alpha(X,Y) \neq H_\alpha(Y) + H_\alpha(X/Y) \qquad (46)$$

and from (40), (42), (43) and (44) we have in general for $\alpha > 1/2$, $\alpha \neq 1$

$$H_\alpha(X) + H_\alpha(Y/X) \neq H_\alpha(Y) + H_\alpha(X/Y)$$

(38) follows from this and (45) and (46).

4. SOME MATHEMATICAL DIFFICULTIES IN ADOPTING $H_\alpha(X)$ AS A MEASURE OF α- UNCERTAINTY.

Renyi's α-entropy given by (1) for discrete scheme is positive, finite and invariant under the transformations of coordinate systems. It will be shown that, in comparison to it, its extension $H_\alpha(X)$ given by (4) for a continuous scheme has none of these basic properties.

4.1 <u>$H_\alpha(X)$ may be negative</u>

Considering
$$f(x) = \frac{3x^2}{a^3} \qquad 0 \leqslant x \leqslant a$$
$$= 0 \qquad \text{elsewhere}$$

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \int_0^a (3x^2/a^3)^\alpha \, dx$$

$$= \frac{1}{1-\alpha} \log \left(\frac{3^\alpha a^{-\alpha+1}}{2\alpha+1} \right) \tag{47}$$

So from (47), for $0 < \alpha < 1$,

$$H_\alpha(X) > 0 \qquad \text{if} \quad a > \left(\frac{2\alpha+1}{3^\alpha} \right)^{1/(1-\alpha)}$$

$$= 0 \qquad \text{if} \quad a = \left(\frac{2\alpha+1}{3^\alpha} \right)^{1/(1-\alpha)}$$

$$< 0 \qquad \text{if} \quad a < \left(\frac{2\alpha+1}{3^\alpha} \right)^{1/(1-\alpha)}$$

and, for $\alpha > 1$,

$$H_\alpha(X) > 0 \qquad \text{if} \quad a < \left(\frac{2\alpha+1}{3^\alpha} \right)^{1/(1-\alpha)}$$

$$= 0 \qquad \text{if} \quad a = \left(\frac{2\alpha+1}{3^\alpha} \right)^{1/(1-\alpha)}$$

$$< 0 \qquad \text{if} \quad a > \left(\frac{2\alpha+1}{3^\alpha} \right)^{1/(1-\alpha)}$$

4.2 $H_\alpha(X)$ may be infinitely large

Let X, having the probability density $f(x)$, takes a continuum of values in $[a, b]$ so that

$$P(l < X < m) = \int_l^m f(x)\,dx = F(m) - F(l)$$

Divide (a, b) into n+1 non-overlapping subintervals $(a, a_1]$, $(a_1, a_2]$, ..., $(a_n, b]$ and let $a_1 - a = \triangle a_1, \ldots,$ $b - a_n = \triangle a_{n+1}$. So

$$P(a < X \leqslant a_1) = \int_a^{a_1} f(x)\,dx = F(a_1) - F(a) = p_1 \triangle a_1$$

$$P(a_n < X \leqslant b) = \int_{a_n}^{b} f(x)\,dx = F(b) - F(a_n) = p_{n+1} \triangle a_{n+1}$$

Define another random variable X_d (say) assuming only the discrete set of values $[a_1, \ldots\ldots, a_n, b]$ with respective probabilities

$$\left[p_1 \triangle a_1, \ldots\ldots\ldots\ldots, p_{n+1} \triangle a_{n+1}\right] \qquad (48)$$

such that $\sum_{i=1}^{n+1} p_i \triangle a_i = F(b) - F(a) = 1$. By (48),

Renyi's α-entropy for the variable X_d is given by

$$(1/1-\alpha) \log \sum_{i=1}^{n+1} p_i^\alpha (\triangle a_i)^\alpha \qquad (49)$$

Under the conditions it is reasonable to expect (49) approach, in the limit as $n \to \infty$, to the α-entropy associated with the continuous variable X. Letting $\triangle a_1 = \ldots\ldots = \triangle a_{n+1} = \triangle x$ in (49), it reduces to

$$(1/1-\alpha) \log \left(\sum_{i=1}^{n+1} p_i^{\alpha} \right) \triangle x + (1/1-\alpha) \log (\triangle x)^{\alpha-1}$$

$$= -\log \triangle x + (1/1-\alpha) \log \left(\sum_{i=1}^{n+1} p_i^{\alpha} \right) \triangle x \qquad (50)$$

In the limit, as $n \to \infty$ i.e., $\triangle x \to 0$, (50) reduces to

$$-\underset{\triangle x \to 0}{\mathrm{Lt}} \log \triangle x + (1/1-\alpha) \log \int_a^b f^{\alpha}(x)\, dx \qquad (51)$$

So in the limit, when an infinite no. of infinitesimal subintervals are considered, $H_\alpha(X)$ given by (51) becomes infinitely large for $\alpha > 0$.

4.3 $H_\alpha(X)$ <u>may not be invariant under a transformation of coordinate systems</u>.

Let the variable X be transformed into a new variable Y by a continuous, monotonic and 1-1 transformation $Y = g(X)$. The density function $\rho(y)$ of the variable Y in terms of $f(x)$ is given by

$$\rho(y) = f(x)\, |dx/dy|$$

So the α-entropy associated with Y is

$$H_\alpha(Y) = (1/1-\alpha) \log \int_{-\infty}^{\infty} f^\alpha(x) \, |dx/dy|^{\alpha-1} \, dx \qquad (52)$$

The value of the integral on the R.H.S. of (52) depends upon $|dx/dy|^{\alpha-1}$ or $|dy/dx|^{1-\alpha}$, a function of x.

For example, letting $Y = AX + B$, we have

$$H_\alpha(Y) = H_\alpha(X) + \log |A|$$

5. DEFINITIONS AND PROPERTIES OF α-INFORMATIONS

As shown in section 4 $H_\alpha(X)$ does not possess the three basic properties of being positive, finite and invariant under even linear transformations, so this obscures its significance. But , nevertheless, it retains its importance , for the α-informations defined below depend on the difference of two α-entropies. These α-informations possess the above basic properties and may be termed as informations of order α processed in the channel.

We may define α-information, $\alpha > 0$, $\alpha \neq 1$, conveyed about X by Y as

$$I_\alpha(X ;Y) = H_\alpha(X) - H_\alpha(X/Y) \qquad (53)$$

and α- information, $\alpha > 0$, $\alpha \neq 1$, conveyed about Y by X as

$$I_\alpha(Y; X) = H_\alpha(Y) - H_\alpha(Y/X) \qquad (54)$$

Note that in general $I_\alpha(X; Y) \neq I_\alpha(Y; X)$.

[by (38)]

We may term $I_\alpha(X;Y)$ and $I_\alpha(Y;X)$ as α-informations of type I and type II respectively.

In the limit, as $\alpha \to 1$, both $I_\alpha(X;Y)$ and $I_\alpha(Y;X)$ tend to $I_1(X;Y)$, transinformation for continuous channels.

5.1 <u>$I_\alpha(X;Y)$ and $I_\alpha(Y;X)$ are always positive</u>

This follows from (20) and (21).

<u>Note</u> that $I_\alpha(X;Y)$ and $I_\alpha(Y;X)$ are zero only when X and Y are independent.

5.2 <u>$I_\alpha(X;Y)$ and $I_\alpha(Y;X)$ are generally finite</u>

When from (50) the expression for $H_\alpha(X)$ was approximated as a limiting case of the corresponding expression for the discrete case there appeared in (51) a term

$$-\operatorname*{Lt}_{\Delta x \to 0} \log \Delta x \tag{55}$$

that led us to say that $H_\alpha(X)$ can be infinitely large. So when an expression for $I_\alpha(X;Y)$ or $I_\alpha(Y;X)$ is approximated in the same way as a limiting case of the corresponding expression for the discrete case there would appear two same terms of the type given in (55), one with the positive sign and the other with the negative sign ; they would cancel each other and thus the resulting expression for $I_\alpha(X;Y)$ or $I_\alpha(Y;X)$ would generally be finite.

5.2 $I_\alpha(X\,;\,Y)$ and $I_\alpha(Y\,;\,X)$ are invariant under linear transformations.

In the original coordinate system let X and Y have the densities $f(x)$, $f(y)$, $f(x,y)$, $f(y/x)$ and $f(x/y)$ subjected to

$$X' = a\,X + b \quad , \quad Y' = c\,Y + d \tag{56}$$

let the above densities be respectively transformed to $\phi(x')$, $\phi(y')$, $\phi(x',y')$, $\phi(y'/x')$, $\phi(x'/y')$, then we know that

$$f(x,y) = \phi(x',y')\left|\frac{\partial(x',y')}{\partial(x,y)}\right| = \phi(x',y')|a\;c|,$$

$$f(x) = \phi(x')|a| \text{ etc.}$$

So

$$H_\alpha(X) - H_\alpha(X/Y) = \frac{1}{1-\alpha}\log\int_{-\infty}^{\infty} f^{\alpha}(x)dx - \frac{1}{1-\alpha}\int_{-\infty}^{\infty} f(y)(\log\int_{-\infty}^{\infty} f^{\alpha}(x/y)dx)dy$$

$$= \frac{1}{1-\alpha}\log\int_{-\infty}^{\infty}\phi^{\alpha}(x')|a|^{\alpha-1}dx' - \frac{1}{1-\alpha}\int_{-\infty}^{\infty}\phi(y')(\log\int_{-\infty}^{\infty}\phi^{\alpha}(x'/y')|a|^{\alpha-1}dx')dy'$$

$$= \frac{1}{1-\alpha}\log\int_{-\infty}^{\infty}\phi^{\alpha}(x')dx' + \frac{1}{1-\alpha}\log|a|^{\alpha-1}$$

$$- \frac{1}{1-\alpha}\log|a|^{\alpha-1} - \frac{1}{1-\alpha}\int_{-\infty}^{\infty}\phi(y')(\log\int_{-\infty}^{\infty}\phi^{\alpha}(x'/y')dx)dy'$$

$$= H_\alpha(X') - H_\alpha(X'/Y') \tag{57}$$

Similarly we can prove that, subject to linear transformation defined in (56)

$$H_\alpha(Y) - H_\alpha(Y/X) = H_\alpha(Y') - H_\alpha(Y'/X') \qquad (58)$$

6. α -CAPACITIES

Two types of α-informations defined in section 5 in a natural way give rise to two types of α-capacities, $C_\alpha^{(1)}$ and $C_\alpha^{(2)}$. So α- capacity of type I may be defined as the maximum of α-information of type I and α-capacity of type II as the maximum of α-information of type II over the whole class of admissible input probability density functions. Symbolically

$$C_\alpha^{(1)} = \max_{f(x)} I_\alpha(X; Y) = \max \left[H_\alpha(X) - H_\alpha(X/Y)\right] \qquad (59)$$

and

$$C_\alpha^{(2)} = \max_{f(x)} I_\alpha(Y; X) = \max \left[H_\alpha(Y) - H_\alpha(Y/X)\right] \qquad (60)$$

In the limit, as $\alpha \to 1$, both $C_\alpha^{(1)}$ and $C_\alpha^{(2)}$ tend to C_1.

Since in general $I_\alpha(X; Y) \neq I_\alpha(Y ; X)$, so in general $C_\alpha^{(1)} \neq C_\alpha^{(2)}$.

7. ALTERNATIVE DEFINITION OF CONDITIONAL α-ENTROPY AND VERIFICATION OF RESULTS

As in the discrete case, conditional α-entropy $H_\alpha(Y/X)$ can be defined in another possible way. Instead of taking the

mean value of $H_\alpha(Y/x)$ the way we did for deriving (9) from (8) we can also consider again the strictly monotonic and continuous function $g_\alpha(x)$ given by (5) for ''information contained'' variable $H_\alpha(Y/x)$ and take the generalised mean value of $H_\alpha(Y/x)$ w.r.t. $g_\alpha(x)$ as we did implicitly for getting (8) or explicitly for getting the extension (6) of Renyi's α-entropy. So if we do this, the conditional α-entropy would be

$$H_\alpha(Y/X) = g_\alpha^{-1}\left[\int_{-\infty}^{\infty} f(x)\left(2^{(1-\alpha)H_\alpha(Y/x)}\right)dx\right] \qquad (61)$$

Using (8), (61) reduces to

$$H_\alpha(Y/X) = g_\alpha^{-1}\left[\int_{-\infty}^{\infty} f(x)\left(2^{\log_2 \int_{-\infty}^{\infty} f^\alpha(y/x)dy}\right)dx\right]$$

$$=(1/1-\alpha)\log_2 \int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} f^\alpha(y/x)dy\right)dx \qquad (62)$$

$$\alpha > 0,\ \alpha \neq 1$$

Based on the same procedure as adopted for deriving (62) we have

$$H_\alpha(X/Y) =(1/1-\alpha)\log_2\int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{\infty} f^\alpha(x/y)\,dx\right)dy \qquad (63)$$

$$\alpha > 0,\ \alpha \neq 1 .$$

In the limit, as $\alpha \to 1$, (62), and (63) tend to the corresponding Shannon's definitions.

But by Jensen's inequality

$$\int_{-\infty}^{\infty} f(x) \left(\log_2 \int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right)dx \leq \log_2 \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right) dx$$

so for $0 < \alpha < 1$,

$$(1/1-\alpha) \int_{-\infty}^{\infty} f(x)\left(\log_2 \int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right)dx \leq (1-\alpha)\log_2 \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right)dx \quad (64)$$

and for $\alpha > 1$,

$$(1/1-\alpha) \int_{-\infty}^{\infty} f(x)\left(\log_2 \int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right)dx \geq (1/1-\alpha)\log_2 \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^{\infty} f^{\alpha}(y/x)dy\right)dx \quad \ldots. (65)$$

By (64) and (65) conditional α-entropy given by (9) is less for $0< \alpha < 1$ and more for $\alpha > 1$, $\alpha \neq \infty$ than that given by (62). Similar interpretation can be had for (10) by comparing it with (63) by Jensen's inequality. Based on this it follows that α-informations defined by (53) and (54) would be less for $0 < \alpha < 1$ and more for $\alpha > 1$ if we measure $H_{\alpha}(Y/X)$ and $H_{\alpha}(X/Y)$ by (62) and (63) in place of (9) and (10).

Results, which involve specifically $H_{\alpha}(Y/X)$ or $H_{\alpha}(X/Y)$ and given by the following equations, are also true if we use the possible alternative definitions of $H_{\alpha}(Y/X)$ and $H_{\alpha}(X/Y)$

given respectively by (62) and (63) :

(20), (21), (30), (33), (34),(35), (37), (38), (43), (57), (58).

In fact all the results that involve either $H_\alpha(Y/X)$ or $H_\alpha(X/Y)$ and are developed in earlier sections, except those given by (25) and (31), are true for these alternative definitions.

Note the generalisation of results developed in section 3 for any finite number of univariate continuous schemes are not taken up in this part since these generalisations also follow from the generalisations for any finite number of multivariate continuous schemes of the corresponding results for two multivariate continuous schemes. These corresponding results and their generalisations are taken up in part II.

It may also be noted that the results developed in sections 2.4, 2.5 and 3.4 of chapter 2 can also be extended to the continuous case.

PART - II

1. INTRODUCTION AND α-ENTROPY ASSOCIATED WITH A RANDOM VECTOR

Suppose the input of a channel is an n-dimensional continuous random vector $\tilde{X} = [X_1, \ldots, X_n]$ having the probability density $f(\tilde{x})$ positive everywhere in its domain of definition R_n - an n-dimensional rectangle. Following the procedure adopted for getting (6) of part I if we consider the

strictly monotonic and continuous parametric vector function

$$g_{\alpha}(\bar{x}) = 2^{(1-\alpha)\,\bar{x}} \qquad \alpha > 0,\ \alpha \neq 1 \qquad (1)$$

and take the Kolmogorov-Nagumo generalised mean value of $\log_2(1/f(\bar{x}))$ w.r.t. (1) we shall obtain α-entropy associated with $\bar{X}$. Hence

$$H_{\alpha}(\bar{X}) = g_{\alpha}^{-1}\left[\int_{R_n} f(\bar{x})\, 2^{(1-\alpha)\log_2(1/f(\bar{x}))}\, d\bar{x}\right]$$

$$= (1/1-\alpha)\log_2 \int_{R_n} f^{\alpha}(\bar{x})\, d\bar{x} \qquad \alpha > 0,\ \alpha \neq 1 \quad (2)$$

Note that either by adopting Shannon's convention that $0 \times \infty = 0$ or by considering the evaluation of $H_{\alpha}(\bar{X})$ in the extended real number system the definition (2) can be made valid even when $f(\bar{x})$ vanishes at the most over a set of measure zero.

In the limit, as $\alpha \to 1$, (2) tends to the corresponding Shannon's definition.

By taking (2) as the basic measure of entropy some basic concepts of multidimensional continuous memoryless channels have been generalised. All the results developed below tend in the limit, as $\alpha \to 1$, to the available results. The results needed are :

(i) ''Jensen's inequality for multiple integrals'', page 21, Morrey (1966). Since the interpretation of the result given there demands that the means of the components of the vector, of whose convex function the theorem deals with, should exist, so our results would be confined only to such distributions.

(ii) '' Holder inequality for multiple integrals'', page 21, Beckenbach and Bellman (1965). There will be strict equality in the results of Holder inequality when the underlying multidimensional schemes are independent.

It may be noted that the brief proofs of the results in sections 2.2 and 2.3 are given in spite of the fact that these results are straight forward generalisations of the corresponding results dealt with in part I because of the following reasons:

(i) the proof for the result in section 2.2 is applicable whether the underlying channel is with memory or without memory whereas the proof given for (20), part I, is applicable to channels without memory.

(ii) the proofs of the results given by (60), (61), (62), (63)and (64) are based respectively on the lines of the proofs of (24), (30), (31), (33) and (50); if the complete proofs of the former results are resorted to, the presentation would be much lengthier that what it is; so to save space, to make the presentation easily assimilable and more or less independent of its two dimensional counterpart dealt with in part I, the brief but complete proofs of the latter results are given whereas the proofs of the former results are simply outlined.

It may also be noted that the operations dealt with below are contingent upon the existence of the underlying integrals. From section 2.2 onwards, base of the logs will be assumed to be e.

2. α- ENTROPIES AND SOME RELATIONS BETWEEN THEM

In 2.1, α-entropies associated with a multidimensional continuous channel are defined and in 2.2, 2.3 and 2.4 some relations between them are developed.

2.1 α-Entropies

Let the input of the channel be an n-dimensional random vector $\check{X} = [X_1, \ldots, X_n]$ with the density function $f(\check{x})$ defined over an n-dimensional rectangle R_n and the output be an m-dimensional random vector $\check{Y} = [Y_1, \ldots, Y_m]$ with the density function $f(\check{y})$ defined over an m-dimensional rectangle R_m. Taken jointly both input and output would be defined over the subspaces of the product space R_{m+n}. There are five α-entropies associated with the product space R_{m+n}, representing the channel.

(i) $H_\alpha(\check{X})$, defined for the input $\check{X}$. It is given by (2).

(ii) $H_\alpha(\check{Y})$, defined for the output $\check{Y}$. Based on the procedure adopted for getting (3) it is given by

$$H_\alpha(\check{Y}) = (1/1-\alpha) \log_2 \int_{R_m} f^\alpha(\check{y})\, d\check{y} \quad \alpha > 0,\ \alpha \neq 1 \quad (3)$$

(iii) $H_\alpha(\check{X}, \check{Y})$, defined for the system as a whole.

Based on the procedure adopted for getting (2) it is given by

$$H_\alpha(\tilde{X}, \tilde{Y}) = (1/1-\alpha) \log_2 \int_{R_{m+n}} f^\alpha(\tilde{x}, \tilde{y}) \, d\tilde{x} \, d\tilde{y} \tag{4}$$

(iv) $H_\alpha(\tilde{Y}/\tilde{X})$, defined about the output $\tilde{Y}$ when it is known that $\tilde{X}$ is transmitted and may be termed as conditional α-entropy. It may be defined as such :

Based on the procedure adopted for getting (2), conditional α-entropy calculated on the assumption that the event $\tilde{x}$ of $\tilde{X}$ has occurred is

$$H_\alpha(\tilde{Y}/\tilde{x}) = (1/1-\alpha)\log_2 \int_{R_m} f^\alpha(\tilde{y}/\tilde{x}) \, d\tilde{y} \quad , \quad \alpha > 0, \alpha \neq 1 \tag{5}$$

To have a measure for $H_\alpha(\tilde{Y}/\tilde{X})$ it is quite natural to take the expected value of (5); so

$$H_\alpha(\tilde{Y}/\tilde{X}) = (1/1-\alpha) \int_{R_n} f(\tilde{x}) \left[\log_2 \int_{R_m} f^\alpha(\tilde{y}/\tilde{x}) \, d\tilde{y}\right] d\tilde{x} \tag{6}$$

$$\alpha > 0, \alpha \neq 1 .$$

(v) $H_\alpha(\tilde{X}/\tilde{Y})$, defined about the input $\tilde{X}$ when it is known that $\tilde{Y}$ is received and be termed as α-equivocation. Based on the procedure adopted for getting (6) it is given by

$$H_\alpha(\tilde{X}/\tilde{Y}) = (1/1-\alpha) \int_{R_m} f(\tilde{y}) \left[\log_2 \int_{R_n} f^\alpha(\tilde{x}/\tilde{y}) \, d\tilde{x}\right] d\tilde{y} \tag{7}$$

$$\alpha > 0, \alpha \neq 1.$$

It can be easily verified that in the limit, as $\alpha \to 1$, (3), (4), (6 and (7) tend to the corresponding Shannon's definitions.

Since

$$\underset{\alpha \to \infty}{\mathrm{Lt}} \left[\int_{R_n} f(\tilde{x})\, f^{\alpha-1}(\tilde{x})\ d\tilde{x} \right]^{1/\alpha-1} = \max\ f(\tilde{x})$$

so

$$\underset{\alpha \to \infty}{\mathrm{Lt}} \quad H_\alpha(\tilde{X}) = -\log_2 \max\ f(\tilde{x})$$

$$\underset{\alpha \to \infty}{\mathrm{Lt}} \quad H_\alpha(\tilde{X}, \tilde{Y}) = -\log_2 \max_{\tilde{x},\tilde{y}}\ f(\tilde{x}, \tilde{y})$$

$$\underset{\alpha \to \infty}{\mathrm{Lt}} \ H_\alpha(\tilde{Y}/\tilde{X}) = -\int_{R_n} f(\tilde{x}) \left[\log_2 \max_{\tilde{y}}\ f(\tilde{y}/\tilde{x})\right] d\tilde{x}$$

2.2 Relation Between $H_\alpha(\tilde{Y}/\tilde{X})$ And $H_\alpha(\tilde{Y})$

Case when $0 < \alpha < 1$

Since $\log \int_{R_m} f^\alpha(\tilde{y}/\tilde{x})\, d\tilde{y}$ is a concave function of $\tilde{x}$, so by Jensen's inequality for multiple integrals.

$$\int_{R_n} f(\tilde{x}) \left[\log f^\alpha(\tilde{y}/\tilde{x})\, d\tilde{y}\right] d\tilde{x} \leq \log \int_{R_n}\int_{R_m} f(\tilde{x})\, f^\alpha(\tilde{y}/\tilde{x})\, d\tilde{x}\, d\tilde{y}$$

$$= \log \int_{R_m} \left[\int_{R_n} f^{1-\alpha}(\tilde{x})\, f^\alpha(\tilde{x},\tilde{y})\, d\tilde{x}\right] d\tilde{y} \qquad \ldots (8)$$

In $\int_{R_n} f^{1-\alpha}(\tilde{x})\, f^\alpha(\tilde{x},\tilde{y})\, d\tilde{x}$, let $1/p = \alpha$ and $1/q = 1-\alpha$ so that $1/p + 1/q = 1$ and applying Hölder's inequality for multiple integrals for $p > 1$, page 21, Beckenbach and

Bellman (1965), we have

$$\int_{R_n} f^{1-\alpha}(\bar{x})\, f^{\alpha}(\bar{x},\bar{y})\, d\bar{x} \leq \left(\int_{R_n} f(\bar{x})\, d\bar{x}\right)^{1-\alpha} \left(\int_{R_n} f(\bar{x},\bar{y})\, d\bar{x}\right)^{\alpha}$$

$$= f^{\alpha}(\bar{y}) \qquad (9)$$

Substituting (9) in (8) and dividing the result by $(1/1-\alpha)$, a positive number, we have

$$H_{\alpha}(\bar{Y}/\bar{X}) \leq H_{\alpha}(\bar{Y}) \qquad 0 < \alpha < 1 \qquad (10)$$

Case when $\alpha > 1$

Since

$$\int_{R_m} f^{\alpha}(\bar{y}/\bar{x})\, d\bar{y} \leq \left(\int_{R_m} f(\bar{y}/\bar{x})\, d\bar{y}\right)^{\alpha} = 1$$

So

$$\int_{R_n} f(\bar{x})\left(\log \int_{R_m} f^{\alpha}(\bar{y}/\bar{x})d\bar{y}\right)d\bar{x} \geq \int_{R_n} f(\bar{x}) \left[\left(\int_{R_m} f^{\alpha}(\bar{y}/\bar{x})d\bar{y}\right) \log \left(\int_{R_m} f^{\alpha}(\bar{y}/\bar{x})d\bar{y}\right)\right] d\bar{x} \qquad \ldots(11)$$

Since $\left(\int_{R_m} f^{\alpha}(\bar{y}/\bar{x})d\bar{y}\right) \log \left(\int_{R_m} f^{\alpha}(\bar{y}/\bar{x})d\bar{y}\right)$ is a convex function of $\bar{x}$, so using Jensen's inequality for multiple integrals (11) reduces to

$$\int_{R_n} f(\bar{x}) \left(\log \int_{R_m} f^{\alpha}(\bar{y}/\bar{x})d\bar{y}\right)d\bar{x} \geq \left[\int_{R_n} f(\bar{x})\left(\int_{R_m} f^{\alpha}(\bar{y}/\bar{x})d\bar{y}\right)d\bar{x}\right] \log \left[\int_{R_n} f(\bar{x}) \left(\int_{R_m} f^{\alpha}(\bar{y}/\bar{x})d\bar{y}\right)d\bar{x}\right] \qquad (12)$$

But $\int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})\, d\tilde{y} \leq 1$; let $\int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})\, d\tilde{y} = \beta_{\tilde{x}}$ (say) ≤ 1.

Let $\beta = \min_{\tilde{x}} \beta_{\tilde{x}}$; using this (12) reduces to

$$\int_{R_n} f(\tilde{x})(\log \int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y})d\tilde{x} \geq \beta \log \int_{R_m} (\int_{R_n} f^{1-\alpha}(\tilde{x})f^{\alpha}(\tilde{x},\tilde{y})d\tilde{x})d\tilde{y} \qquad \ldots(13)$$

In $\int_{R_n} f^{1-\alpha}(\tilde{x})\, f^{\alpha}(\tilde{x},\tilde{y})\, d\tilde{x}$ let $1/p = \alpha$ and $1/q = 1-\alpha$

so that $1/p + 1/q = 1$ and applying Hölder's inequality for multiple integrals for $p < 1$, we have

$$\int_{R_n} f^{1-\alpha}(\tilde{x})f^{\alpha}(\tilde{x},\tilde{y})d\tilde{x} \geq (\int_{R_n} f(\tilde{x})\, d\tilde{x})^{1-\alpha} (\int_{R_n} f(\tilde{x},\tilde{y})\, d\tilde{x})^{\alpha} = f^{\alpha}(\tilde{y}) \qquad \ldots (14)$$

Using (14) in (13) and multiplying the result so obtained by $(1/1-\alpha)$, a negative number, we shall have

$$H_{\alpha}(\tilde{Y}/\tilde{X}) \leq \beta\, H_{\alpha}(\tilde{Y}) \qquad (15)$$

Since $\beta \leq 1$, so from (15) we have

$$H_{\alpha}(\tilde{Y}/\tilde{X}) \leq H_{\alpha}(\tilde{Y}) \qquad \alpha > 1, \qquad (16)$$

Combining (10) and (16) we have

$$H_{\alpha}(\tilde{Y}/\tilde{X}) \leq H_{\alpha}(\tilde{Y}) \qquad \alpha > 0,\ \alpha \neq 1, \qquad (17)$$

In the limit, as $\alpha \to 1$, (17) tends to $H_1(\tilde{Y}/\tilde{X}) \leq H_1(\tilde{Y})$ as already known as Shannon's inequality, so combining this

symbolically with (17) we shall have

$$H_\alpha(\tilde{Y}/\tilde{X}) \leq H_\alpha(\tilde{Y}) \qquad \alpha > 0 \qquad (18)$$

Similarly we can prove

$$H_\alpha(\tilde{X}/\tilde{Y}) \leq H_\alpha(\tilde{X}) \qquad \alpha > 0 \qquad (19)$$

2.3 Relations Between $H_\alpha(\tilde{X},\tilde{Y})$, $H_\alpha(\tilde{X})$ And $H_\alpha(\tilde{Y}/\tilde{X})$

Two infinite sets of inequalities, each member of either tending in the limit, as $\alpha \to 1$, to the known equality $H_1(\tilde{X},\tilde{Y}) = H_1(\tilde{X}) + H_1(\tilde{Y}/\tilde{X})$, and two bounds for $H_\alpha(\tilde{X},\tilde{Y})$, one upper for $0 < \alpha < 1$ and one lower for $\alpha > 1$, are developed.

Case (i) for $0 < \alpha < 1$

$$\log \int_{R_{m+n}} f^\alpha(\tilde{x},\tilde{y})d\tilde{x}\, d\tilde{y} = \log \int_{R_n} f(\tilde{x})^{\alpha-1/p}\left[\int_{R_m} f(\tilde{x})^{1/p}\; f^\alpha(\tilde{y}/\tilde{x})d\tilde{y}\right] d\tilde{x} \quad (20)$$

where $0 < p < 1$. Choosing q subject to $1/p + 1/q = 1$ and applying Hölder inequality for multiple integrals for this value of the parameter p we have

$$\log \int_{R_{m+n}} f^\alpha(\tilde{x},\tilde{y})d\tilde{x}\, d\tilde{y} \geq \log \left[\int_{R_n} (f^{\alpha-1/p}(\tilde{x}))^q\, d\tilde{x}\right]^{1/q}$$

$$+ \log \left[\int_{R_n} \left(\int_{R_m} f^{1/p}(\tilde{x})\; f^\alpha(\tilde{y}/\tilde{x})d\tilde{y}\right)^p d\tilde{x}\right]^{1/p}$$

$$\ldots(21)$$

But by Jensen's inequality for multiple integrals

$$\log \int_{R_n} f(\tilde{x})\left(\int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y}\right)^{p} d\tilde{x} \geqslant p \int_{R_n} f(\tilde{x})\left(\log \int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y}\right)d\tilde{x} \quad \ldots . (22)$$

Using (22), (21) reduces to

$$\log \int_{R_{m+n}} f^{\alpha}(\tilde{x},\tilde{y})d\tilde{x}\, d\tilde{y} \geqslant \frac{1}{q} \log \int_{R_n} f^{(\alpha-1/p)q}(\tilde{x})d\tilde{x} + \int_{R_n} f(\tilde{x})\left(\log \int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y}\right)d\tilde{x} \quad \ldots (23)$$

Multiplying (23) by $(1/1-\alpha)$, a negative number, we have

$$H_{\alpha}(\tilde{X},\tilde{Y}) \geqslant H_{(\alpha-1/p)q}(\tilde{X}) + H_{\alpha}(\tilde{Y}/\tilde{X}) \qquad (24)$$

for $0<\alpha<1$, $0<p<1$ and $1/p + 1/q = 1$

Under the conditions for which (24) is valid, $(\alpha - 1/p)q$ lies between 1 and ∞. So (24) gives an infinite set of inequalities connecting $H_{\alpha}(\tilde{X}, \tilde{Y})$ and $H_{\alpha}(\tilde{Y}/\tilde{X})$ for $0<\alpha<1$ to $H_{\alpha}(\tilde{X})$ for $\alpha > 1$. It can be easily verified that in the limit, as $\alpha \to 1$, every member of (24) tends to the known equality $H_1(\tilde{X},\tilde{Y}) = H_1(\tilde{X}) + H_1(\tilde{Y}/\tilde{X})$.

Case (ii) for $0<\alpha<1$

Choosing in (20), $p > 1$ and q subject to $1/p + 1/q = 1$ and applying Hölder inequality for multiple integrals for this value of the parameter p we have

$$\log \int_{R_{m+n}} f^{\alpha}(\tilde{x},\tilde{y})d\tilde{x}\,d\tilde{y} \leq \log \left[\int_{R_n} (f^{\alpha-1/p}(\tilde{x}))^q d\tilde{x}\right]^{1/q} +$$

$$+ \log \left[\int_{R_n} \left(\int_{R_m} f^{1/p}(\tilde{x}) f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y}\right)^p d\tilde{x}\right]^{1/p} \quad (25)$$

Multiplying (25) by $(1/1-\alpha)$, a positive number, we have

$$H_{\alpha}(\tilde{X},\tilde{Y}) \leq H_{(\alpha-1/p)q}(\tilde{X}) + \frac{1}{(1-\alpha)p} \log \int_{R_n} f(\tilde{x}) \left[\int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y}\right]^p d\tilde{x} \quad (26)$$

for $0 < \alpha < 1$, $p > 1$ and $1/p + 1/q = 1$

Taking limits of (26) as $p \to \infty$ we have

$$\underset{p \to \infty}{\mathrm{Lt}} \; H_{(\alpha-1/p)q}(\tilde{X}) = H_{\alpha}(\tilde{X}) \quad (27)$$

$$\underset{p \to \infty}{\mathrm{Lt.}} \; \frac{1}{(1-\alpha)p} \log \int_{R_n} f(\tilde{x}) \left[\int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})\,d\tilde{y}\right]^p d\tilde{x}$$

$$= \underset{p \to \infty}{\mathrm{Lt}} \int_{R_n} \frac{f(\tilde{x}) \left(\int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y}\right)^p}{\int_{R_n} f(\tilde{x}) \left(\int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y}\right)^p d\tilde{x}} \left(\frac{1}{1-\alpha} \log \int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})d\tilde{y}\right) d\tilde{x} \quad \ldots (28)$$

The R.H.S. of (28) is the weighted mean of

$\frac{1}{1-\alpha} \log \int_{R_m} f^{\alpha}(\tilde{y}/\tilde{x})\,d\tilde{y}$ i.e. $H_{\alpha}(\tilde{Y}/\tilde{x})$, so

$$\text{L.H.S. of (28)} \leq \max_{\tilde{x}} H_{\alpha}(\tilde{Y}/\tilde{x}) \quad (29)$$

Using (27) and (29), (28) in the limit, as $p \to \infty$, reduces to

$$H_\alpha(\tilde{X}, \tilde{Y}) \leq H_\alpha(\tilde{X}) + \max_{\tilde{x}} H_\alpha(\tilde{Y}/\tilde{x}) \tag{30}$$

(30) gives an upper bound for $H_\alpha(\tilde{X}, \tilde{Y})$ for $0 < \alpha < 1$.

<u>Case (i) for $\alpha > 1$</u>

Since in deriving (23) no condition on α was made, so multiplying (23) by $(1/1-\alpha)$, a negative number, we have

$$H_\alpha(\tilde{X},\tilde{Y}) \leq H_{(\alpha-1/p)q}(\tilde{X}) + H_\alpha(\tilde{Y}/\tilde{X}) \tag{31}$$

for $\alpha > 1$, $\cancel{\alpha < p < \infty}$, $0 < p < 1$ and $1/p + 1/q = 1$

Under the conditions for which (31) is valid $(\alpha-1/p)q < 1$; so (31) gives an infinite set of inequalities connecting $H_\alpha(\tilde{X},\tilde{Y})$ and $H_\alpha(\tilde{Y}/\tilde{X})$ for $\alpha > 1$, to $H_\alpha(\tilde{X})$ for $0 < \alpha < 1$. It can be easily verified that in the limit, as $\alpha \to 1$, every member of (31) tends to $H_1(\tilde{X}, \tilde{Y}) = H_1(\tilde{X}) + H_1(\tilde{Y}/\tilde{X})$.

<u>Case (ii) for $\alpha > 1$</u>

Multiplying (25) by $(1/1-\alpha)$, a negative number, we shall have

$$H_\alpha(\tilde{X},\tilde{Y}) \geq H_{(\alpha-1/p)q}(\tilde{X}) + \frac{1}{(1-\alpha)p} \log \int_{R_n} f(\tilde{x}) \left(\int_{R_m} f^\alpha(\tilde{y}/\tilde{x}) d\tilde{y} \right)^p d\tilde{x} \qquad \ldots \quad (32)$$

In the limit, as $p \to \infty$, (32) tends to

$$H_\alpha(\tilde{X},\tilde{Y}) \geqslant H_\alpha(\tilde{X}) + \min_{\tilde{x}} H_\alpha(\tilde{Y}/\tilde{x}) \qquad (33)$$

(33) gives a lower bound for $H_\alpha(\tilde{X},\tilde{Y})$ for $\alpha > 1$.

Similarly we can prove that for $0 < \alpha < 1$,

$$H_\alpha(\tilde{X},\tilde{Y}) \geqslant H_{(\alpha-1/p)q}(\tilde{Y}) + H_\alpha(\tilde{X}/\tilde{Y}) \qquad 0< p < 1, \ 1/p + 1/q = 1$$

$$H_\alpha(\tilde{X},\tilde{Y}) \leqslant H_\alpha(\tilde{Y}) + \max_{y} H_\alpha(\tilde{X}/\tilde{y})$$

and for $\alpha > 1$,

$$H_\alpha(\tilde{X},\tilde{Y}) \leqslant H_{(\alpha-1/p)q}(\tilde{Y}) + H_\alpha(\tilde{X}/\tilde{Y}) \qquad 0< p < 1, \ 1/p + 1/q = 1$$

$$H_\alpha(\tilde{X},\tilde{Y}) \geqslant H_\alpha(\tilde{Y}) + \min_{y} H_\alpha(\tilde{X}/\tilde{y})$$

2.4 A Study of Equality

It will be proved that if the joint distribution of $\tilde{X}$ and $\tilde{Y}$ is gaussian then, for $\alpha > 0$,

$$H_\alpha(\tilde{X},\tilde{Y}) = H_\alpha(\tilde{X}) + H_\alpha(\tilde{Y}/\tilde{X}) = H_\alpha(\tilde{Y}) + H_\alpha(\tilde{X}/\tilde{Y})$$

An example that this is not, in general, true is also indicated in the end.

Below, throughout this part, a single integral sign will stand for a multiple integral and the number of variables for which it stands is understood from the dimension of the variable

vector.

Let $\tilde{Z}$ be the random vector describing the joint behaviour of $\tilde{X}$ and $\tilde{Y}$ i.e. $\tilde{Z} = [\tilde{X}, \tilde{Y}] = [X_1, \ldots, X_n, Y_1, \ldots, Y_m]$. Without loss of generality we may assume that $E(\tilde{X}) = \tilde{0}$ and $E(\tilde{Y}) = \tilde{0}$.

With this the probability density of $\tilde{Z}$ is given by

$$f(\tilde{z}) = \frac{1}{(2\pi)^{(m+n)/2} \, |C|^{1/2}} \exp. \left[- \frac{1}{2} (\tilde{z}' C^{-1} \tilde{z}) \right]$$

where $\tilde{z}'$ is the transpose of $\tilde{z}$ and the matrix C is given by

$$C = \begin{bmatrix} A & D \\ D' & B \end{bmatrix}$$

In the partitioned structure of C, the matrix D' is the transpose of D and A, B and D are respectively given by

$$A = [a_{ij}] , \quad a_{ij} = \int_{-\infty}^{\infty} x_i \, x_j \, f(\tilde{x}) \, d\tilde{x} \qquad i,j = 1,2,\ldots,n$$

$$B = [b_{ij}] , \quad b_{ij} = \int_{-\infty}^{\infty} y_i \, y_j \, f(\tilde{y}) \, d\tilde{y} \qquad i,j = 1,2,\ldots,m$$

$$D = [d_{ij}] , \quad d_{ij} = \int_{-\infty}^{\infty} x_i \, y_j \, f(\tilde{z}) \, d\tilde{z} \qquad i = 1,\ldots,n, \; j = 1,\ldots,m$$

The marginal densities for $\tilde{X}$ and $\tilde{Y}$ are respectively

given by

$$f(\tilde{x}) = \frac{1}{(2\pi)^{n/2} |A|^{1/2}} \exp. \left[-\frac{1}{2} (\tilde{x}' A^{-1} \tilde{x}) \right]$$

and

$$f(\tilde{y}) = \frac{1}{(2\pi)^{m/2} |B|^{1/2}} \exp. \left[-\frac{1}{2} (\tilde{y}' B^{-1} \tilde{y}) \right]$$

The conditional density of $\tilde{Y}$ given a value $\tilde{x}$ of $\tilde{X}$ is given by

$$f(\tilde{y}/\tilde{x}) = \frac{f(\tilde{x},\tilde{y})}{f(\tilde{x})} = \frac{f(\tilde{z})}{f(\tilde{x})} = \frac{|A|^{1/2}}{(2\pi)^{m/2} |C|^{1/2}}$$

$$\exp \left[-\frac{1}{2} (\tilde{z}' C^{-1} \tilde{z} - \tilde{x}' A^{-1} \tilde{x}) \right] \quad (34)$$

Let $C^{-1} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$ so that

$$C\,C^{-1} = \begin{bmatrix} A & D \\ D' & B \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} I_n & 0_{n\times m} \\ 0_{m\times n} & I_m \end{bmatrix}$$

i.e.

$$A\,S_{11} + D\,S_{21} = I_n \tag{35}$$

$$A\,S_{12} + D\,S_{22} = 0_{n \times m} \tag{36}$$

$$D'\,S_{12} + B\,S_{22} = I_m \tag{37}$$

$$D'\,S_{11} + B\,S_{21} = 0_{m\times n} \tag{38}$$

Solving (35) and (36) we have

$$A^{-1} = S_{11} - S_{12}\,S_{22}^{-1}\,S_{21}$$

This would make

$$\tilde{z}'C^{-1}\tilde{z} - \tilde{x}'A^{-1}\tilde{x} = \begin{bmatrix}\tilde{x}' & \tilde{y}'\end{bmatrix}\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}\begin{bmatrix}\tilde{x} \\ \tilde{y}\end{bmatrix} - \tilde{x}'(S_{11}-S_{12}S_{22}^{-1}S_{21})\tilde{x}$$

$$= (\tilde{y} + S_{22}^{-1}\,S_{21}\tilde{x})'\,S_{22}\,(\tilde{y} + S_{22}^{-1}\,S_{21}\,\tilde{x}) \tag{39}$$

Hence, for $\alpha > 0$, $\alpha \neq 1$,

$$H_\alpha(\tilde{X},\tilde{Y}) = \frac{1}{1-\alpha}\log\int_{-\infty}^{\infty}\left(\frac{1}{(2\pi)^{\frac{m+n}{2}}\,|C|^{1/2}}\exp(-1/2\begin{bmatrix}\tilde{x} & \tilde{y}\end{bmatrix}'\,C^{-1}\begin{bmatrix}\tilde{x} & \tilde{y}\end{bmatrix})\right)^{\alpha} d\tilde{x}d\tilde{y}$$

$$\ldots (40)$$

Putting in (40) $\sqrt{\alpha}\,\tilde{X} = \tilde{P}$ and $\sqrt{\alpha}\,\tilde{Y} = \tilde{Q}$, it will be simplified to

$$H_\alpha(\tilde{X},\tilde{Y}) = \log (2\pi)^{\frac{m+n}{2}}\,|C|^{1/2} - \frac{m+n}{2(1-\alpha)}\log\alpha \tag{41}$$

$$H_\alpha(\tilde{X}) = \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} \left(\frac{1}{(2\pi)^{n/2} |A|^{1/2}} \exp. -\frac{1}{2} (\tilde{x}' A^{-1} \tilde{x}) \right)^\alpha d\tilde{x} \quad .. (42)$$

Putting in (42) $\sqrt{\alpha}\, \tilde{x} = \tilde{P}$, it will be simplified to

$$H_\alpha(\tilde{X}) = \log (2\pi)^{n/2} |A|^{1/2} - \frac{n}{2(1-\alpha)} \log \alpha \quad (43)$$

Similarly

$$H_\alpha(\tilde{Y}) = \log (2\pi)^{m/2} |B|^{1/2} - \frac{m}{2(1-\alpha)} \log \alpha \quad (44)$$

Using (34) and (39) we have

$$H_\alpha(\tilde{Y}/\tilde{X}) = \frac{1}{1-\alpha} \int_{-\infty}^{\infty} f(\tilde{x}) \left[\log \int_{-\infty}^{\infty} \left(\frac{|A|^{1/2}}{(2\pi)^{m/2} |C|^{1/2}} \exp. -\frac{1}{2} (\tilde{y} + S_{22}^{-1} S_{21} \tilde{x})' S_{22} (\tilde{y} + S_{22}^{-1} S_{21} \tilde{x}) \right)^\alpha d\tilde{y} \right] d\tilde{x}$$

$$= \frac{\alpha}{1-\alpha} \log \frac{|A|^{1/2}}{(2\pi)^{\frac{m}{2}} |C|^{\frac{1}{2}}} + \frac{1}{1-\alpha} \int_{-\infty}^{\infty} f(\tilde{x})$$

$$\left(\log \int_{-\infty}^{\infty} \exp. \left(\frac{\alpha}{2} (\tilde{y} + S_{22}^{-1} S_{21} \tilde{x})' S_{22} (\tilde{y} + S_{22}^{-1} S_{21} \tilde{x}) \right) d\tilde{y} \right) d\tilde{x} \quad ... (45)$$

Putting $\sqrt{\alpha}\, (\tilde{y} + S_{22}^{-1} S_{21} \tilde{x}) = \tilde{T}$ in the second term of (45), it will reduce to

$$H_\alpha(\check{Y}/\check{X}) = \frac{\alpha}{1-\alpha} \log \frac{|A|^{1/2}}{(2\pi)^{m/2}\,|C|^{1/2}} - \frac{m}{2(1-\alpha)} \log \alpha$$

$$- \frac{1}{1-\alpha} \log (2\pi)^{m/2} \,|S_{22}|^{1/2} \tag{46}$$

By solving (36) and (37) we have $S_{22}^{-1} = B - D' A^{-1} D$.

Letting $L = \begin{bmatrix} I & -(A^{-1})' D \\ 0 & I \end{bmatrix}$, We have

$$L' C L = \begin{bmatrix} I & 0 \\ -D'A^{-1} & I \end{bmatrix} \begin{bmatrix} A & D \\ D' & B \end{bmatrix} \begin{bmatrix} I & -(A^{-1})' D \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} A & 0 \\ 0 & -D'A^{-1}D + B \end{bmatrix} \quad \text{Since } A^{-1} \text{ is symmetric}$$

so $|C| = |A|\;|B - D' A^{-1} D|$

Hence $$|S_{22}|^{-1/2} = |B - D'A^{-1}D|^{1/2} = \frac{|C|^{1/2}}{|A|^{1/2}} \tag{47}$$

Using (47),(46) reduces to

$$H_\alpha(\check{Y}/\check{X}) = \log \frac{(2\pi)^{m/2}\,|C|^{1/2}}{|A|^{1/2}} - \frac{m}{2(1-\alpha)} \log \alpha \tag{48}$$

Similarly we can prove

$$H_\alpha(\tilde{X}/\tilde{Y}) = \log \frac{(2\pi)^{n/2} \, |C|^{1/2}}{|B|^{1/2}} - \frac{n}{2(1-\alpha)} \log \alpha \qquad (49)$$

Using (41), (43) and (48) we have,

$$H_\alpha(\tilde{X},\tilde{Y}) = H_\alpha(\tilde{X}) + H_\alpha(\tilde{Y}/\tilde{X}) \qquad \alpha > 0,\ \alpha \neq 1 \quad (50)$$

and using (41), (44) and (49) we have

$$H_\alpha(\tilde{X},\tilde{Y}) = H_\alpha(\tilde{Y}) + H_\alpha(\tilde{X}/\tilde{Y}) \qquad \alpha > 0,\ \alpha \neq 1 \quad (51)$$

(50) and (51) gives for $\alpha > 0$, $\alpha \neq 1$,

$$H_\alpha(\tilde{X},\tilde{Y}) = H_\alpha(\tilde{X}) + H_\alpha(\tilde{Y}/\tilde{X}) = H_\alpha(\tilde{Y}) + H_\alpha(\tilde{X}/\tilde{Y})$$

From this and the fact, that this is true for $\alpha = 1$, the required result follows.

But in general, for $\alpha > 0$, $\alpha \neq 1$,

$$H_\alpha(\tilde{X}, \tilde{Y}) \neq H_\alpha(\tilde{X}) + H_\alpha(\tilde{Y}/\tilde{X})$$

It can be shown by considering a simple example. An example in this regard is the bivariate scheme given by

$$f(x,y) = 2 \quad \text{for} \quad \begin{cases} 0 < x < 1 \\ 0 < y < x \end{cases}$$
$$= 0 \quad \text{outside}$$

3. GENERALISATIONS

In this section various results developed in section 2 are generalised for any finite number of multivariate continuous schemes.

Consider r spaces R_{n_1}, $R_{n_2}, \ldots,$ R_{n_r} of dimensions $n_1, n_2, \ldots, n_r$ with respective continuous variable points

$$\tilde{x}_1 = \left[x_{11}, x_{12} \ldots\ldots\ldots, x_{1n_1}\right]$$

$$\tilde{x}_2 = \left[x_{21}, x_{22}, \ldots\ldots\ldots, x_{2n_2}\right]$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\tilde{x}_r = \left[x_{r1}, x_{r2}, \ldots\ldots, x_{r\,n_r}\right]$$

having the respective density functions $f(\tilde{x}_1), \ldots, f(\tilde{x}_r)$ so that

$$\int_{R_{n_1}} f(\tilde{x}_1)\, d\tilde{x}_1 = 1, \ \ldots\ldots\ldots, \ \int_{R_{n_r}} f(\tilde{x}_r)\, d\tilde{x}_r = 1$$

The product space of these r spaces is of dimension $n_1 + n_2 + \ldots\ldots + n_r$ and it may be symbolically denoted by $R_{n_1+n_2+\ldots+n_r}$. Let the variable point

$$\tilde{z} = \left[\tilde{x}_1, \ldots, \tilde{x}_r\right] = \left[x_{11}, \ldots\ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2}; \ldots\ldots, x_{r1}, \ldots\ldots, x_{rn_r}\right]$$

for the product space have $f(\tilde{z})$ as its density function so that $\int_{R_{n_1+\ldots+n_r}} f(\tilde{z})\, d\tilde{z} = 1$. In this product space the following are a few generalisations, along with either complete proofs or sufficient indications in that direction, of the various results developed in sections 2.2, 2.3 and 2.4.

(i) The inequality in (18) will get generalised to inequalities of the form

$$H_\alpha(\tilde{X}_r / \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{r-1}) \leq H_\alpha(\tilde{X}_r / \tilde{X}_2, \tilde{X}_3, \ldots, \tilde{X}_{r-1}) \qquad (52)$$

$$\alpha > 0$$

Proof of (52) :

On the assumption that an event $\left[\tilde{X}_2, \ldots, \tilde{X}_{r-1}\right]$ of the product space $R_{n_2+n_3+\ldots+n_{r-1}}$ has occurred, let the density function of the events belonging to $R_{n_1+n_r}$ be given by

$$f(\tilde{x}_1, \tilde{x}_r / \tilde{x}_2, \tilde{x}_3, \ldots, \tilde{x}_{r-1}) = g(\tilde{x}_1, \tilde{x}_r)$$

and the schemes given by the variable points $\tilde{X}_1$ and $\tilde{X}_r$ become the schemes given by the variables $\tilde{X}_1' = \left[X'_{11}, .., X'_{1n_1}\right]$ and $\tilde{X}_r' = \left[X'_{r1}, \ldots, X'_{rn_r}\right]$ with respective density functions

$$f(\tilde{x}_1 / \tilde{x}_2, \ldots, \tilde{x}_{r-1}) = g(\tilde{x}_1)$$

and

$$f(\tilde{x}_r / \tilde{x}_2, \ldots, \tilde{x}_{r-1}) = g(\tilde{x}_r)$$

By (18) for $\alpha > 0$, $\alpha \neq 1$ we have

$$H_\alpha(\tilde{X}_r' / \tilde{X}_1') \leq H_\alpha(\tilde{X}_r') \qquad (53)$$

But

$$H_\alpha(\tilde{X}_r' / \tilde{X}_1') = \frac{1}{1-\alpha} \int_{R_{n_1}} g(\tilde{x}_1) \left(\log \int_{R_{n_r}} g^\alpha(\tilde{x}_r / \tilde{x}_1)\, d\tilde{x}_r\right) d\tilde{x}_1 \qquad (54)$$

and

$$g(\tilde{x}_r/\tilde{x}_1) = \frac{g(\tilde{x}_r,\tilde{x}_1)}{g(\tilde{x}_1)} = \frac{f(\tilde{x}_1,\tilde{x}_r/\tilde{x}_2,\tilde{x}_3,\ldots,\tilde{x}_{r-1})}{f(\tilde{x}_1/\tilde{x}_2,\tilde{x}_3,\ldots,\tilde{x}_{r-1})} =$$

$$= f(\tilde{x}_r/\tilde{x}_1,\ldots,\tilde{x}_{r-1}) \qquad (55)$$

Using (55), (54) reduces to

$$H_\alpha(\tilde{X}_r'/\tilde{X}_1') = \frac{1}{1-\alpha}\int_{R_{n_1}} f(\tilde{x}_1/\tilde{x}_2,\ldots,\tilde{x}_{r-1})(\log\int_{R_{n_r}} f^\alpha(\tilde{x}_r/\tilde{x}_1,\ldots,\tilde{x}_{r-1})d\tilde{x}_r)d\tilde{x}_1 \qquad \ldots (56)$$

Also

$$H_\alpha(\tilde{X}_r') = \frac{1}{1-\alpha}\log\int_{R_{n_r}} g^\alpha(\tilde{x}_r)\, d\tilde{x}_r$$

$$= \frac{1}{1-\alpha}\log\int_{R_{n_r}} f^\alpha(\tilde{x}_r/\tilde{x}_2,\ldots,\tilde{x}_{r-1})\, d\tilde{x}_r \qquad (57)$$

Using (56) and (57), (53) reduces to

$$\frac{1}{1-\alpha}\int_{R_{n_1}} f(\tilde{x}_1/\tilde{x}_2,\ldots,\tilde{x}_{r-1})(\log\int_{R_{n_r}} f^\alpha(\tilde{x}_r/\tilde{x}_1,\ldots,\tilde{x}_{r-1})d\tilde{x}_r)\,d\tilde{x}_1$$

$$\leq \frac{1}{1-\alpha}\log\int_{R_{n_r}} f^\alpha(\tilde{x}_r/\tilde{x}_2,\ldots,\tilde{x}_{r-1})d\tilde{x}_r \qquad (58)$$

Multiplying (58) by $f(\tilde{x}_2,\ldots,\tilde{x}_{r-1})$ and integrating over $R_{n_2+\ldots+n_{r-1}}$ we shall have

$$H_\alpha(\tilde{X}_r/\tilde{X}_1,\tilde{X}_2,\ldots,\tilde{X}_{r-1}) \leq H_\alpha(\tilde{X}_r/\tilde{X}_2,\ldots,\tilde{X}_{r-1}) \qquad (59)$$

$$\alpha > 0,\ \alpha \neq 1,$$

In the limit, as $\alpha \to 1$, (59) tends to the known result

$H_1(\tilde{X}_r/\tilde{X}_1,\tilde{X}_2,\ldots,\tilde{X}_{r-1}) \leq H_1(\tilde{X}_r/\tilde{X}_2,\ldots,\tilde{X}_{r-1})$. Combining this symbolically with (59) we shall have (52).

(ii) For $0 < \alpha < 1$, the inequalities in (24) will get generalised to inequalities of the form

$$H_\alpha(\tilde{X}_1,\ldots,\tilde{X}_r) \geq H_{(\alpha-1/p_1)p_1'}(\tilde{X}_1) + H_{(\alpha-1/p_2)p_2'}(\tilde{X}_2/\tilde{X}_1) + \ldots$$

$$+ H_{(\alpha-1/p_{r-1})p_{r-1}'}(\tilde{X}_{r-1}/\tilde{X}_1,\tilde{X}_2,\ldots,\tilde{X}_{r-2})$$

$$+ H_\alpha(\tilde{X}_r/\tilde{X}_1, \tilde{X}_2,\ldots,\tilde{X}_{r-1}) \qquad (60)$$

where $0 < p_1,\ldots, p_{r-1} < 0$ and $1/p_1 + 1/p_1' = 1,\ldots\ldots,$ $\dfrac{1}{p_{r-1}} + \dfrac{1}{p'_{r-1}} = 1$ and the inequality in (30) to inequalities of the form

$$H_\alpha(\tilde{X}_1,\ldots,\tilde{X}_r) \leq H_\alpha(\tilde{X}_1) + \max_{\tilde{x}_1} H_\alpha(\tilde{X}_2/\tilde{x}_1) + \max_{\tilde{x}_1,\tilde{x}_2} H_\alpha(\tilde{X}_3/\tilde{x}_1,\tilde{x}_2)$$

$$+\ldots+ \max_{\tilde{x}_1,\ldots,\tilde{x}_{r-1}} H_\alpha(\tilde{X}_r/\tilde{x}_1,\ldots,\tilde{x}_{r-1}) \qquad (61)$$

(iii) For $\alpha > 1$, the inequalities in (31) will get generalised to inequalities of the form

$$H_\alpha(\tilde{X}_1,\ldots,\tilde{X}_r) \leqslant H_{(\alpha-1/p_1)p_1'}(\tilde{X}_1) + H_{(\alpha-1/p_2)p_2'}(\tilde{X}_2/\tilde{X}_1)+\ldots$$

$$+ H_{(\alpha-1/p_{r-1})p_{r-1}'}(\tilde{X}_{r-1}/\tilde{X}_1,\tilde{X}_2,\ldots,\tilde{X}_{r-2})$$

$$+ H_\alpha(\tilde{X}_r/\tilde{X}_1,\ \tilde{X}_2,\ldots,\tilde{X}_{r-1}) \qquad (62)$$

where $0< p_1,\ldots,p_{r-1}< 1$ and $\frac{1}{p_1} + \frac{1}{p_1'} = 1,\ldots,\frac{1}{p_{r-1}} + \frac{1}{p_{r-1}'} = 1$

and the inequality in (33) to inequalities of the form

$$H_\alpha(\tilde{X}_1,\ldots,\tilde{X}_r) \geqslant H_\alpha(\tilde{X}_1) + \min_{\tilde{x}_1} H_\alpha(\tilde{X}_2/\tilde{x}_1)+ \min_{\tilde{x}_1,\tilde{x}_2} H_\alpha(\tilde{X}_3/\tilde{x}_1,\tilde{x}_2)$$

$$+\ldots+ \min_{\tilde{x}_1,\ldots,\tilde{x}_{r-1}} H_\alpha(\tilde{X}_r/\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_{r-1}) \qquad (63)$$

(iv) when the vector variables $\tilde{X}_1,\ldots,\tilde{X}_r$ are normally distributed, the equality in (50) will get generalised to equalities of the form

$$H_\alpha(\tilde{X}_1,\ldots,\tilde{X}_r)=H_\alpha(\tilde{X}_1)+ H_\alpha(\tilde{X}_2/\tilde{X}_1)+\ldots+H_\alpha(\tilde{X}_r/\tilde{X}_1,\ldots,\tilde{X}_{r-1}) \qquad (64)$$

<u>Outline of the Proof For (62)</u> :

Applying the procedure followed to get (24), we shall have

$$H_\alpha(\tilde{X}_1,\ldots,\tilde{X}_r) \geqslant H_{(\alpha-1/p_1)p_1'}(\tilde{X}_1) + H_\alpha(\tilde{X}_2,\tilde{X}_3,\ldots,\tilde{X}_r/\tilde{X}_1) \qquad (65)$$

where $0 < p_1 < 1$ and $(1/p_1) + (1/p_1') = 1$.

$$H_\alpha(\tilde{X}_2,\ldots,\tilde{X}_r/\tilde{X}_1) = \frac{1}{1-\alpha}\int_{R_{n_1}} f(\tilde{x}_1)\Big(\log \int_{R_{n_2+\ldots+n_r}} f^\alpha(\tilde{x}_2,\ldots,\tilde{x}_r/\tilde{x}_1)d\tilde{x}_2\ldots d\tilde{x}_r\Big)\, d\tilde{x}_1 \qquad (66)$$

But

$$\log \int_{R_{n_2+\ldots+n_r}} f^\alpha(\tilde{x}_2,\ldots,\tilde{x}_r/\tilde{x}_1)d\tilde{x}_2\ldots d\tilde{x}_r$$

$$= \log \int_{R_{n_2}} f^{\alpha-1/p_2}(\tilde{x}_2/\tilde{x}_1)\left[\int_{R_{n_3+\ldots+n_r}} f^{1/p_2}(\tilde{x}_2/\tilde{x}_1) f^\alpha(\tilde{x}_3,\ldots,\tilde{x}_r/\tilde{x}_1,\tilde{x}_2)\, d\tilde{x}_3\ldots d\tilde{x}_r\right] d\tilde{x}_2 \qquad \ldots (67)$$

where $0 < p_2 < 1$. Choosing p_2' so that $(1/p_2)+(1/p_2')=1$, applying Hölder inequality and following the procedure adopted to get (24), we shall have from (67)

$$H_\alpha(\tilde{X}_2,\ldots,\tilde{X}_r/\tilde{x}_1) \geqslant H_{(\alpha-1/p_2)p_2'}(\tilde{X}_2/\tilde{x}_1) + \frac{1}{1-\alpha}\int_{R_{n_2}} f(\tilde{x}_2/\tilde{x}_1)\Big(\log \int_{R_{n_3+\ldots+n_r}} f^\alpha(\tilde{x}_3,\ldots,\tilde{x}_r/\tilde{x}_1,\tilde{x}_2)d\tilde{x}_3\ldots d\tilde{x}_r\Big)d\tilde{x}_2 \qquad \ldots.(68)$$

Multiplying (68) by $f(\tilde{x}_1)$, integrating over R_{n_1} and making use of (66) , we shall have

$$H_\alpha(\tilde{X}_2,\ldots,\tilde{X}_r/\tilde{X}_1) \geqslant H_{(\alpha-1/p_2)p_2'}(\tilde{X}_2/\tilde{X}_1)$$

$$+ H_\alpha(\tilde{X}_3,\ldots,\tilde{X}_r/\tilde{X}_1,\ \tilde{X}_2) \qquad (69)$$

Continuing the operation of getting (69) for another r-3 times, we shall have (60).

(61), (62), (63) and (64) can similarly be proved by repeating r-1 times the procedure adopted for getting respectively (30), (31), (33) and (50).

3.1 A Corollary

If the vector variables $\tilde{X}_1, \ldots, \tilde{X}_r$ are normally distributed then for $\alpha > 0$

$$H_\alpha(\tilde{X}_1,\ldots,\tilde{X}_r) \leqslant H_\alpha(\tilde{X}_1) + H_\alpha(\tilde{X}_2) + \ldots + H_\alpha(\tilde{X}_r) \qquad (70)$$

Proof of (70) follows from (64) by using inequalities of the form (52).

From (70) in particular for a vector $\tilde{X} = [X_1,\ldots,X_n]$ distributed normally we shall have

$$H_\alpha(\tilde{X}) \leqslant H_\alpha(X_1) + H_\alpha(X_2) + \ldots + H_\alpha(X_n) \qquad (71)$$

It may be noted that in the next chapter it will be shown that (71) is not true in general.

4. α- INFORMATIONS AND THEIR PROPERTIES

As in part I for a univariate case, it can be shown that $H_\alpha(\tilde{X})$ does not have the three basic properties of being finite, positive and invariant under linear transformations.

In spite of this it has importance since the α-informations defined below depend on the difference of two such α-entropies.

We may define α-information, $\alpha > 0$, $\alpha \neq 1$, conveyed about $\tilde{X}$ by $\tilde{Y}$ as

$$I_\alpha(\tilde{X};\tilde{Y}) = H_\alpha(\tilde{X}) - H_\alpha(\tilde{X}/\tilde{Y}) \tag{72}$$

and α-information, $\alpha > 0$, $\alpha \neq 1$, conveyed about $\tilde{Y}$ by $\tilde{X}$ as

$$I_\alpha(\tilde{Y};\tilde{X}) = H_\alpha(\tilde{Y}) - H_\alpha(\tilde{Y}/\tilde{X}) \tag{73}$$

Note that in general $I_\alpha(\tilde{X};\tilde{Y}) \neq I_\alpha(\tilde{Y};\tilde{X})$

We may term $I_\alpha(\tilde{X};\tilde{Y})$ and $I_\alpha(\tilde{Y};\tilde{X})$ as α-informations of type I and type II respectively.

In the limit, as $\alpha \to 1$, both $I_\alpha(\tilde{X};\tilde{Y})$ and $I_\alpha(\tilde{Y};\tilde{X})$ tend to $I_1(\tilde{X};\tilde{Y})$, transinformation for such channels.

4.1 Positiveness of $I_\alpha(\tilde{X};\tilde{Y})$ and $I_\alpha(\tilde{Y};\tilde{X})$

This follows from (19) and (18).

$I_\alpha(\tilde{X};\tilde{Y}) = I_\alpha(\tilde{Y};\tilde{X}) = 0$ if $\tilde{X}$ and $\tilde{Y}$ are independent.

4.2 Finiteness of $I_\alpha(\tilde{X};\tilde{Y})$ and $I_\alpha(\tilde{Y};\tilde{X})$

As shown in section (4.2) of part I when for a univariate continuous scheme described by a random variable $\tilde{X}$, $H_\alpha(\tilde{X})$ was approximated as a limiting case of the corresponding expression for the discrete case there appeared a term

$$-\operatorname*{Lt}_{\Delta x \to 0} \log \Delta x$$

that let us to say that $H_\alpha(X)$ can be infinitely large, so also if we approximate $H_\alpha(\tilde{X})$ as a limiting case of the corresponding expression when $\tilde{X}$ takes only a finite number of values there would appear a term

$$-\underset{\Delta x_1, \ldots\ldots, \Delta x_n \to 0}{\mathrm{Lt}} \log \Delta\tilde{x} \tag{74}$$

where $\Delta\tilde{x} = \Delta x_1 \ldots\ldots \Delta x_n$, that can lead us to say that $H_\alpha(\tilde{X})$ can be infinitely large. But when an expression for $I_\alpha(\tilde{X}\,;\,\tilde{Y})$ or $I_\alpha(\tilde{Y};\,\tilde{X})$ is approximated in the same way there would appear two same terms of the type given by (74), one with the positive sign and the other with the negative sign; they would cancel each other and thus the resulting expression for $I_\alpha(\tilde{X}\,;\,\tilde{Y})$ or $I_\alpha(\tilde{Y}\,;\,\tilde{X})$ would generally be finite.

4.3. Invariance of $I_\alpha(\tilde{X}\,;\,\tilde{Y})$ and $I_\alpha(\tilde{Y}\,;\,\tilde{X})$

Under a non-singular linear transformation both $I_\alpha(\tilde{X}\,;\,\tilde{Y})$ and $I_\alpha(\tilde{Y};\,\tilde{X})$ are invariant. Let $\tilde{X}$ and $\tilde{Y}$ be transformed respectively to $\tilde{X}'$ and $\tilde{Y}'$ by the non-singular transformations

$$\tilde{X}' = A\,\tilde{X} \quad \text{and} \quad \tilde{Y}' = B\,\tilde{Y} \tag{75}$$

Subjected to these transformations we shall have

$$H_\alpha(\tilde{X}') = H_\alpha(\tilde{X}) + \log|A| \tag{76}$$

and

$$H_\alpha(\tilde{X}'/\tilde{Y}') = H_\alpha(\tilde{X}/\tilde{Y}) + \log|A| \tag{77}$$

Using (76) and (77) we have

$$H_\alpha(\check{X}') - H_\alpha(\check{X}'/\check{Y}') = H_\alpha(\check{X}) - H_\alpha(\check{X}/\check{Y}) \tag{78}$$

Similarly subjected to (75) we shall have

$$H_\alpha(\check{Y}') - H_\alpha(\check{Y}'/\check{X}') = H_\alpha(\check{Y}) - H_\alpha(\check{Y}/\check{X}) \tag{79}$$

4.4 A Corollary

α-entropy of an ensemble has the same numerical value when expressed in terms of either its frequency sampling values or time sampling values.

Proof: In Appendix VI Goldman (1953) it has been shown that the transformation between the time and frequency sampling values can be made a linear measure preserving transformation and we know that under any such transformation the Jacobian of transformation is 1. So by (76) the result follows.

5. SECOND POSSIBLE DEFINITION OF CONDITIONAL α-ENTROPY AND VERIFICATIONS

Another possible way of defining conditional α-entropy $H_\alpha(\check{Y}/\check{X})$ is the following :

Instead of taking the mean value of $H_\alpha(\check{Y}/\check{x})$ the way we did for getting (6) from (5), we may consider again the strictly monotonic and continuous vector function $g_\alpha(\check{x}) = 2^{(1-\alpha)\check{x}}$, $\alpha > 0$, $\alpha \neq 1$, and take the generalised mean of the 'information contained' variable $H_\alpha(\check{Y}/\check{x})$ w.r.t $g_\alpha(\check{x})$ as we did for getting (2). If we resort to

this we shall have

$$H_\alpha(\tilde{Y}/\tilde{X}) = g_\alpha^{-1}\left(\int_{R_n} f(\tilde{x})\, 2^{(1-\alpha)\, H_\alpha(\tilde{Y}/\tilde{x})}\, d\tilde{x}\right) \tag{80}$$

Using (5), (80) reduces to

$$H_\alpha(\tilde{Y}/\tilde{X}) = \frac{1}{1-\alpha} \log_2 \int_{R_n} f(\tilde{x}) \left(\int_{R_m} f^\alpha(\tilde{y}/\tilde{x})\, d\tilde{y}\right) d\tilde{x} \tag{81}$$

$$\alpha > 0,\ \alpha \neq 1,$$

Following the same procedure as used for getting (81) we shall have

$$H_\alpha(\tilde{X}/\tilde{Y}) = \frac{1}{1-\alpha} \log_2 \int_{R_m} f(\tilde{y}) \left(\int_{R_n} f^\alpha(\tilde{x}/\tilde{y})\, d\tilde{x}\right) d\tilde{y} \tag{82}$$

$$\alpha > 0,\ \alpha \neq 1$$

In the limit, as $\alpha \to 1$, (81) and (82) tend to the corresponding Shannon's definitions.

Comparisons of α-informations defined on the basis of (6), (7) and (81), (82) can, in the same manner as in, section 7 of part I, be made by using Jensen's inequality for multiple integrals.

Results, which involve specifically $H_\alpha(\tilde{Y}/\tilde{X})$ and $H_\alpha(\tilde{X}/\tilde{Y})$ and given by the following equations, are also if we use the possible alternative definitions of $H_\alpha(\tilde{Y}/\tilde{X})$ and $H_\alpha(\tilde{X}/\tilde{Y})$ given respectively by (81) and (82) :

(18), (19), (30), (33), (46), (49), (50), (51), (52), (61), (63), (64), (78) and (79).

On the whole all the results, except those given

by (24) and (31) and their generalisations given respectively by (60) and (62), are true also for ~~also~~ these alternative definitions of $H_\alpha(\bar{Y}/\bar{X})$ and $H_\alpha(\bar{X}/\bar{Y})$.

Using these alternative definitions all the results enumerated above can be easily proved by simple manipulations of the procedures adopted for proving them by using the definitions given by (6) and (7). For instance, the truth of the result in (18) for this alternative definition of $H_\alpha(\bar{Y}/\bar{X})$ can be argued as such :

For $0 < \alpha < 1$, comparing (8) and the definitions of $H_\alpha(\bar{Y}/\bar{X})$ given by (6) and (81) and retracing the steps used for obtaining (10) from (8), the result follows :

For $\alpha > 0$, by Jensen's inequality for multiple integrals

$$\log \int_{R_n} f(\bar{x}) \left(\int_{R_m} f^\alpha(\bar{y}/\bar{x}) d\bar{y} \right) d\bar{x} \geqslant \int_{R_n} f(\bar{x}) \left(\log \int_{R_m} f^\alpha(\bar{y}/\bar{x}) d\bar{y}\right) d\bar{x}$$

and further retracing the steps used for getting (16) from (11) we shall have the result. Or else

$$\log \int_{R_n}\int_{R_m} f(\bar{x}) f^\alpha(\bar{y}/\bar{x}) d\bar{x}\, d\bar{y} = \log \int_{R_m} f^\alpha(\bar{y}) \left(\int_{R_n} f^{1-\alpha}(\bar{x}) f^\alpha(\bar{x}/\bar{y}) d\bar{x} \right) d\bar{y} \quad \dots (83)$$

By the argument used for getting (15) from (13) we shall have

$$\int_{R_n} f^{1-\alpha}(\bar{x})\, f^\alpha(\bar{x}/\bar{y})\, d\bar{x} \geqslant 1$$

Using (84), (83) reduces to

$$\log \int_{R_n} \int_{R_m} f(\bar{x})\, f^{\alpha}(\bar{y}/\bar{x})\, d\bar{x}\, d\bar{y} \geqslant \log \int_{R_m} f^{\alpha}(\bar{y})\, d\bar{y} \qquad (85)$$

Multiplying (85) by $(1/1-\alpha)$, a negative number, we shall have the result.

CHAPTER - 6

FURTHER ASPECTS OF MULTIDIMENSIONAL CONTINUOUS MEMORYLESS CHANNELS

1. INTRODUCTION

In part II of chapter 5 the results (18) and (19) and the invalidity of

$$H_\alpha(\tilde{X},\tilde{Y}) = H_\alpha(\tilde{X}) + H_\alpha(\tilde{Y}/\tilde{X}) = H_\alpha(\tilde{Y}) + H_\alpha(\tilde{X}/\tilde{Y}), \quad \alpha > 0, \ \alpha \neq 1$$

in general, as shown by an example in section 3.3 of part I of chapter 5 for a bivariate case, made us define two types of α-informations. These are given by (72) and (73) in part II, chapter 5. Two types of α-informations in a natural way give rise to two types of α-capacities. These are defined in section 2. To evaluate α-capacities we need the maximum value of $H_\alpha(\tilde{X})$ or $H_\alpha(\tilde{Y})$. As pointed out in section 4 of part II and proved in section 4.2 of part I for the univariate case, $H_\alpha(\tilde{X})$ can be infinitely large so this makes the maximisation of $H_\alpha(\tilde{X})$ itself a problem. This has been tackled in section 3 for three types of constraints on $\tilde{X}$. In section 4 the invalidity of Shannon's results $H_1(\tilde{X}) \leq H_1(X_1) + \dots + H_1(X_n)$ in general for $\alpha > 0$, $\alpha \neq 1$ is derived. In section 5.1 evaluation of α-informations when both $\tilde{X}$ and $\tilde{Y}$ are Gaussian and in section 5.2 evaluations of α-capacities under two sets of constraints are dealt with. A generalisation of Shannon-Hartley channel-capacity formula is given in section 6. The validity of these results is examined in

section 7 for the alternative definition of conditional α-entropy. All the results tend in the limit as $\alpha \to 1$ to the corresponding available results.

2. <u>α- CAPACITIES</u>

Two types of α-informations defined by (72) and (73) in part II, chapter 5, give rise to two types of α-capacities $\alpha > 0$, $\alpha \neq 1$ i.e. α-capacity of type I as the maximum of α-information of type I and α-capacity of type II as the maximum of α-information of type II. The maximisation is to be done over the whole class of admissible input density functions. Symbolically

$$C_\alpha^{(1)} = \max_{f(\bar{X})} I_\alpha(\bar{X};\, \bar{Y}) = \max_{f(\bar{x})} \left[H_\alpha(\bar{X}) - H_\alpha(\bar{X}/\bar{Y}) \right] \quad (1)$$

and

$$C_\alpha^{(2)} = \max_{f(\bar{x})} I_\alpha(\bar{Y};\, \bar{X}) = \max_{f(\bar{x})} \left[H_\alpha(\bar{Y}) - H_\alpha(\bar{Y}/\bar{X}) \right] \quad (2)$$

In the limit, as $\alpha \to 1$, both $C_\alpha^{(1)}$ and $C_\alpha^{(2)}$ tend to C_1, Shannon's definition of capacity for such channels. Since in general $I_\alpha(\bar{X};\, \bar{Y}) \neq I_\alpha(\bar{Y};\, \bar{X})$, so in general $C_\alpha^{(1)} \neq C_\alpha^{(2)}$.

3. <u>MAXIMISATION OF $H_\alpha(\bar{X})$</u>

Following are the three sets of constraints on $\bar{X}$ for which the problem of maximisation of $H_\alpha(\bar{X})$ is considered:

(i) When the domain of definition R_n of $\bar{X}$ is bounded i.e. $a_1 < X_1 < b_1, \ldots, a_n < X_n < b_n$.

(ii) when the components $X_1, X_2, \ldots, X_n$ of $\tilde{X}$ assume only non-negative values and have specified first moments $a_1, a_2, \ldots, a_n$ ($a_1, a_2, \ldots, a_n > 0$).

(iii) when the components $X_1, \ldots, X_n$ of $\bar{X}$ have respective specified standard deviations $\sigma_1, \ldots, \sigma_n$.

Discussion of Case (i)

The problem is to maximise $\frac{1}{1-\alpha} \log \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f^{\alpha}(\tilde{x})\, dx_1 \ldots dx_n$ i.e.

maximise $\int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f^{\alpha}(\bar{x})\, dx_1 \ldots dx_n$ for $0 < \alpha < 1$

and min. $\int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f^{\alpha}(\bar{x})\, dx_1 \ldots dx_n$ for $\alpha > 1$, i.e.

opt. $\int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f^{\alpha}(\tilde{x}) dx_1 \ldots dx_n$ subject to

$$\int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(\bar{x})\, dx_1 \ldots dx_n = 1 \qquad (3)$$

We know by the technique of Lagrangian multipliers that the optimisation of $\int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f^{\alpha}(\bar{x}) dx_1 \ldots dx_n$ subject to (3) must satisfy

$$\frac{\partial}{\partial f} f^{\alpha} + \lambda \frac{\partial f}{\partial f} = 0$$

i.e. $f = (-\lambda/\alpha)^{1/\alpha - 1}$

But the condition (3) implies that

$$f = \frac{1}{(b_1 - a_1)------(b_n - a_n)}$$

Hence max. $H_\alpha(\tilde{X})$ subject to (3) is

$$\log\left[(b_1 - a_1)\ \ldots.(b_n - a_n)\right]$$

same as the known result for $H_1(\tilde{X})$, Shannon's entropy for $\tilde{X}$.

<u>Discussion of case (ii)</u>

The problem is to maximise

$$\frac{1}{1-\alpha} \log \int_0^\infty \ldots. \int_0^\infty f^\alpha(\tilde{x})\ dx_1 \ldots. dx_n \quad \text{subject to}$$

$$\int_0^\infty \ldots.. \int^\infty f(\tilde{x})\ dx_1 \ldots.. dx_n = 1 \tag{4}$$

and

$$\left.\begin{aligned} \int_0^\infty \ldots \int_0^\infty x_1\ f(\tilde{x})\ dx_1 \ldots. dx_n &= a_1 \\ \int_0^\infty \ldots. \int_0^\infty x_2\ f(\tilde{x})\ dx_1 \ldots dx_n &= a_2 \\ ------------&------ \\ \int_0^\infty \ldots. \int_0^\infty x_n\ f(\tilde{x})\ dx_1 \ldots. dx_n &= a_n \end{aligned}\right\} \tag{5}$$

By the technique of lagrangian multipliers we know that $f(\tilde{x})$ optimising $\int_0^\infty \ldots \int_0^\infty f^\alpha(\tilde{x})\ dx_1 \ldots. dx_n$ subject to (4) and (5), must satisfy

$$\frac{\partial}{\partial f} f^{\alpha} + \mu \frac{\partial f}{\partial f} + \lambda_1 \frac{\partial}{\partial f}(x_1 f) + \text{--------} + \lambda_n \frac{\partial}{\partial f}(x_n f) = 0$$

i.e. $$f(\bar{x}) = \left(\frac{-\mu - \lambda_1 x_1 - \lambda_2 x_2 - \ldots\ldots - \lambda_n x_n}{\alpha} \right)^{\frac{1}{\alpha - 1}} \tag{6}$$

Using (6), (4) reduces to

$$\int_0^{\infty} \ldots \int_0^{\infty} \left(\frac{-\mu - \lambda_1 x_1 - \ldots\ldots - \lambda_n x_n}{\alpha} \right)^{\frac{1}{\alpha - 1}} dx_1 \ldots\ldots dx_n = 1 \tag{7}$$

In (7) for the integrand to be positive we assume μ, λ_1, ..., λ_n to be negative. Integrating (7) first for x_1, then for x_2 and lastly for x_n it will be found that it is convergent iff $\frac{n-1}{n} < \alpha < 1$ and it would ultimately reduce~~s~~ to

$$\frac{(1/\alpha)^{1/\alpha - 1} \quad (-\mu)^{\frac{1}{\alpha-1} + n}}{\lambda_1 \cdots\cdots \lambda_n \left(\frac{1}{\alpha-1} + 1\right)\left(\frac{1}{\alpha-1} + 2\right) \ldots\ldots \left(\frac{1}{\alpha-1} + n\right)} = 1 \tag{8}$$

Similarly, using (6), every member of the set (5) will be convergent iff $\frac{n}{n+1} < \alpha < 1$ and the set itself would respectively reduce to

$$\frac{(-\mu)^{\frac{1}{\alpha-1} + n + 1}}{\lambda_1^2 \lambda_2 \text{-----} \lambda_n \left(\frac{1}{\alpha-1} + 1\right)\left(\frac{1}{\alpha-1} + 2\right) \ldots\ldots \left(\frac{1}{\alpha-1} + n + 1\right)} = a_1(\alpha)^{\frac{1}{\alpha-1}}$$

$$\frac{(-\mu)^{\frac{1}{\alpha-1}+n+1}}{\lambda_1\lambda_2^2\cdots\lambda_n\left(\frac{1}{\alpha-1}+1\right)\left(\frac{1}{\alpha-1}+2\right)\cdots\left(\frac{1}{\alpha-1}+n+1\right)} = a_2(\alpha)^{\frac{1}{\alpha-1}}$$

$$\frac{(-\mu)^{\frac{1}{\alpha-1}+n+1}}{\lambda_1\lambda_2\cdots\lambda_n^2\left(\frac{1}{\alpha-1}+1\right)\left(\frac{1}{\alpha-1}+2\right)\cdots\left(\frac{1}{\alpha-1}+n+1\right)} = a_n(\alpha)^{\frac{1}{\alpha-1}}$$

$$\ldots (9)$$

Dividing (8) by the respective members of (9) we shall have

$$\left.\begin{aligned}
-\frac{\lambda_1}{\mu}\left(\frac{1}{\alpha-1}+n+1\right) &= \frac{1}{a_1}\\
-\frac{\lambda_2}{\mu}\left(\frac{1}{\alpha-1}+n+1\right) &= \frac{1}{a_2}\\
&\cdots\cdots\cdots\cdots\\
-\frac{\lambda_n}{\mu}\left(\frac{1}{\alpha-1}+n+1\right) &= \frac{1}{a_n}
\end{aligned}\right\} \qquad (10)$$

Multiplying the various members of (10) we have

$$(a_1a_2\cdots a_n)\left(\frac{1}{\alpha-1}+n+1\right)^n = \frac{(-\mu)^n}{\lambda_1\lambda_2\cdots\lambda_n} \qquad (11)$$

So using (6) we have

$$\max \; H_\alpha(\tilde{X}) = \frac{1}{1-\alpha} \log \int_0^\infty \cdots \int_0^\infty \left(\frac{-\mu - \lambda_1 x_1 - \cdots - \lambda_n x_n}{\alpha} \right)^{\frac{\alpha}{\alpha-1}} dx_1 \ldots dx_n \qquad \ldots (12)$$

But

$$\int_0^\infty \cdots \int_0^\infty \left(\frac{-\mu - \lambda_1 x_1 - \cdots - \lambda_n x_n}{\alpha} \right)^{\frac{\alpha}{\alpha-1}} dx_1 \ldots dx_n$$

$$= \frac{(1/\alpha)^{\alpha/\alpha-1} \; (-\mu)^{\alpha/\alpha-1+n}}{\lambda_1 \lambda_2 \cdots \lambda_n \left(\frac{\alpha}{\alpha-1}+1\right)\left(\frac{\alpha}{\alpha-1}+2\right)\cdots\left(\frac{\alpha}{\alpha-1}+n\right)}$$

$$= \frac{-\mu}{\alpha} \qquad \frac{\frac{1}{\alpha-1}+1}{\frac{\alpha}{\alpha-1}+n} \qquad \left[\text{By } (8)\right] \qquad (13)$$

Comparing (8) and (11)

$$-\mu = \left(\frac{(a_1 a_2 \ldots a_n)\left(\frac{1}{\alpha-1}+n+1\right)^n (1/\alpha)^{\frac{1}{\alpha-1}}}{\left(\frac{1}{\alpha-1}+1\right)\left(\frac{1}{\alpha-1}+2\right)\cdots\left(\frac{1}{\alpha-1}+n\right)} \right)^{1-\alpha} \qquad (14)$$

Using (14), (13) reduces to

$$\int_0^{\infty}\cdots\int_0^{\infty} \left(\frac{-\mu - \lambda_1 x_1 - \cdots - \lambda_n x_n}{\alpha} \right)^{\frac{\alpha}{\alpha-1}} dx_1 \ldots dx_n$$

$$= \frac{(a_1 \ldots a_n)^{1-\alpha} \; (n(\alpha-1) + \alpha)^{n(1-\alpha)-1} \; \alpha}{\left[\alpha(2\alpha-1) \ldots\ldots\ldots\ldots (n\alpha - n+1)\right]^{1-\alpha}} \tag{15}$$

Using (15), (12) reduces to

$$\max H_\alpha(\tilde{X}) = \log (a_1 \ldots a_n) + n \log \left[n(\alpha-1) + \alpha\right]$$

$$- \frac{1}{1-\alpha} \log \frac{n(\alpha-1)+\alpha}{\alpha} - \log \left[\alpha(2\alpha-1)\ldots(n\alpha-n+1)\right] \tag{16}$$

$$\frac{n}{n+1} < \alpha < 1$$

In the limit, as $\alpha \to 1$, (16) tends to

$$\max.\ H_1(\tilde{X}) = \log (a_1 \ldots a_n) + n$$

$$= \log a_1 e + \log a_2 e + \ldots + \log a_n e$$

as already known.

<u>Discussion of Case (iii)</u>

Supposing $E(X_i) = 0 \quad i = 1,2,\ldots,n$, the problem is to maximise

$$\frac{1}{1-\alpha} \log \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f^{\alpha}(\tilde{x}) dx_1 \ldots dx_n \text{ i.e.}$$

$$\text{opt. } \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f^{\alpha}(\tilde{x}) dx_1 \ldots dx_n \tag{17}$$

subject to

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\bar{x})\, dx_1 \cdots dx_n = 1 \tag{18}$$

and

$$\left.\begin{array}{l} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^2 f(\bar{x})\, dx_1 \cdots dx_n = \sigma_1^2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_n^2 f(\bar{x})\, dx_1 \cdots dx_n = \sigma_n^2 \end{array}\right\} \tag{19}$$

By the technique of Lagrangian multipliers we know that $f(\bar{x})$, optimising $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f^{\alpha}(\bar{x})\, dx_1 \cdots dx_n$ subject to (18) and (19), must satisfy

$$\frac{\partial}{\partial f} f^{\alpha} - \mu \frac{\partial f}{\partial f} - \lambda_1 \frac{\partial}{\partial f}(x_1^2 f) - \cdots\cdots - \lambda_n \frac{\partial}{\partial f}(x_n^2 f) = 0$$

i.e.
$$f(\bar{x}) = \left(\frac{\mu + \lambda_1 x_1^2 + \cdots\cdots + \lambda_n x_n^2}{\alpha} \right)^{\frac{1}{\alpha-1}} \tag{20}$$

Using (20), (18) reduces to

$$2^n (1/\alpha)^{\frac{1}{\alpha-1}} \int_0^{\infty}\cdots\int_0^{\infty} (\mu + \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2)^{\frac{1}{\alpha-1}} dx_1 \cdots dx_n = 1 \tag{21}$$

The integral in (21) is convergent iff $\frac{2}{1-\alpha} - n + 1 > 1$ i.e. $\frac{n-2}{n} < \alpha < 1$. To integrate (21) over x_1 for fixed values of $x_2, x_3, \ldots, x_n$, we first make the substitution

$$(\mu + \lambda_2 x_2^2 + \ldots + \lambda_n x_n^2) + \lambda_1 x_1^2 = t$$

then

$$\frac{\mu + \lambda_2 x_2^2 + \ldots + \lambda_n x_n^2}{t} = 1$$

in the result so obtained. If we simplify further, we shall have

$$2^{n-1} (1/\alpha)^{\frac{1}{\alpha-1}} \frac{1}{\sqrt{\lambda_1}} \beta\left(\frac{\alpha+1}{2(1-\alpha)}, \frac{1}{2}\right) \int_0^\infty \ldots \int_0^\infty (\mu + \lambda_2 x_2^2 + \ldots + \lambda_n x_n^2)^{\frac{\alpha+1}{2(\alpha-1)}} dx_2 \ldots dx_n = 1 \qquad (22)$$

Repeating the above process for $x_2, \ldots, x_n$ we shall ultimately have

$$(1/\alpha)^{1/\alpha-1} \frac{1}{\sqrt{\lambda_1 \cdots \lambda_n}} \mu^{\frac{1}{\alpha-1}+\frac{n}{2}} \beta\left(\frac{1}{1-\alpha} - \frac{1}{2}, \frac{1}{2}\right) \cdots \cdots \beta\left(\frac{1}{1-\alpha} - \frac{n}{2}, \frac{1}{2}\right) = 1 \qquad (23)$$

Using (20) in the respective members of the set of equations (19) and making use of the procedure adopted to get (23) from (21), we shall find that every member of the set (19) is convergent iff $\frac{n}{n+2} < \alpha < 1$ and the members of (19) would respectively reduce to

$$\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} \frac{\mu^{\frac{3\alpha-1}{2(\alpha-1)}+\frac{n+1}{2}}}{\lambda_1^{3/2} \sqrt{\lambda_2 \cdots \lambda_n}} \beta\left(\frac{3\alpha-1}{2(1-\alpha)}, \frac{3}{2}\right) \beta\left(\frac{3\alpha-1}{2(1-\alpha)} - \frac{1}{2}, \frac{1}{2}\right) \ldots \beta\left(\frac{3\alpha-1}{2(1-\alpha)} - \frac{n-1}{2}, \frac{1}{2}\right) = \sigma_1^2$$

$$\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} \frac{\mu^{\frac{3\alpha-1}{2(\alpha-1)} + \frac{n-1}{2}}}{\lambda_2^{3/2}\sqrt{\lambda_1 \lambda_3 \cdots\cdots \lambda_n}} \beta\left(\frac{3\alpha-1}{2(1-\alpha)}, \frac{3}{2}\right) \beta\left(\frac{3\alpha-1}{2(1-\alpha)} - \frac{1}{2}, \frac{1}{2}\right) \cdots\cdots \beta\left(\frac{3\alpha-1}{2(1-\alpha)} - \frac{n-1}{2}, \frac{1}{2}\right) = \sigma_2^2$$

$$(1/\alpha)^{1/\alpha-1} \frac{\mu^{\frac{3\alpha-1}{2(\alpha-1)} + \frac{n-1}{2}}}{\lambda_n^{3/2}\sqrt{\lambda_1 \lambda_2 \cdots\cdots \lambda_{n-1}}} \beta\left(\frac{3\alpha-1}{2(1-\alpha)}, \frac{3}{2}\right) \beta\left(\frac{3\alpha-1}{2(1-\alpha)} - \frac{1}{2}, \frac{1}{2}\right) \cdots\cdots \beta\left(\frac{3\alpha-1}{2(1-\alpha)} - \frac{n-1}{2}, \frac{1}{2}\right) = \sigma_n^2 \quad (24)$$

Dividing (23) with the respective members of (24) we shall have

$$\frac{\lambda_1}{\mu} \frac{n(\alpha-1) + 2\alpha}{1-\alpha} = \frac{1}{\sigma_1^2}$$

-- (25)

$$\frac{\lambda_n}{\mu} \frac{n(\alpha-1) + 2\alpha}{1-\alpha} = \frac{1}{\sigma_n^2}$$

Multiplying the members of (25) we shall have

$$(\sigma_1^2 \cdots\cdots \sigma_n^2)\left(\frac{n(\alpha-1) + 2\alpha}{1-\alpha}\right)^n = \frac{\mu^n}{\lambda_1 \lambda_2 \cdots\cdots \lambda_n} \quad (26)$$

Using (20), (17) reduces to

$$\max.\ H_\alpha(\tilde{X}) = \frac{1}{1-\alpha}\log \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{\mu+\lambda_1 x_1^2+\cdots+\lambda_n x_n^2}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} dx_1\cdots dx_n \quad (27)$$

where μ , λ_1 ,..., λ_n are given by (23) and (24). But applying the procedure adopted for getting (23) from (21), we have

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{\mu+\lambda_1 x_1^2+\cdots+\lambda_n x_n^2}{\alpha}\right)^{\alpha/\alpha-1} dx_1\cdots dx_n$$

$$=(1/\alpha)^{\alpha/\alpha-1}\frac{\mu^{\frac{\alpha}{\alpha-1}+\frac{n}{2}}}{\sqrt{\lambda_1\cdots\lambda_n}}\ \beta\left(\frac{\alpha}{1-\alpha}-\frac{1}{2},\frac{1}{2}\right)\cdots\beta\left(\frac{\alpha}{1-\alpha}-\frac{n}{2},\frac{1}{2}\right) \quad \ldots (28)$$

Using (23), (28) reduces to

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{\mu+\lambda_1 x_1^2+\cdots+\lambda_n x_n^2}{\alpha}\right)^{\alpha/\alpha-1} dx_1\cdots dx_n$$

$$= \frac{\mu}{\alpha}\ \frac{\beta\left(\frac{\alpha}{1-\alpha}-\frac{1}{2},\frac{1}{2}\right)\cdots\beta\left(\frac{\alpha}{1-\alpha}-n/2,\ 1/2\right)}{\beta(1/1-\alpha-1/2,1/2)\cdots\beta(1/1-\alpha-n/2,1/2)} \quad (29)$$

Comparing (23) and (26) we shall have

$$\mu = \left[(1/\alpha)^{2/\alpha-1}\left(\frac{n(\alpha-1)+2\alpha}{1-\alpha}\right)^n(\sigma_1^2\cdots\cdots\sigma_n^2)\right.$$
$$\left.\left(\beta\left(\frac{1}{1-\alpha}-1/2,\ 1/2\right)\cdots\beta\left(\frac{1}{1-\alpha}-\frac{n}{2},\ \frac{1}{2}\right)\right)^2\right]^{\frac{1-\alpha}{2}} \quad (30)$$

Using (30) in (29) and simplifying we shall have

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{\mu+\lambda_1 x_1^2+\cdots+\lambda_n x_n^2}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} dx_1\cdots dx_n$$

$$=\frac{2\alpha(\sigma_1\cdots\sigma_n)^{1-\alpha}\left[n(\alpha-1)+2\alpha\right]^{n\left(\frac{1-\alpha}{2}\right)-1}\left[\beta\left(\frac{1}{1-\alpha}-\frac{1}{2},\frac{1}{2}\right)\cdots\beta\left(\frac{1}{1-\alpha}-\frac{n}{2},\frac{1}{2}\right)\right]^{1-\alpha}}{(1-\alpha)^{n\left(\frac{1-\alpha}{2}\right)}} \quad \ldots\ldots (31)$$

Using (31), (27) reduces to

$$\max H_\alpha(\tilde{X})=\frac{1}{1-\alpha}\log\frac{2\alpha}{n(\alpha-1)+2\alpha}+\log\left[\beta\left(\frac{1}{1-\alpha}-\frac{1}{2},\frac{1}{2}\right)\cdots\beta\left(\frac{1}{1-\alpha}-\frac{n}{2},\frac{1}{2}\right)\right]$$

$$-\frac{1}{2}\log(1-\alpha)^n+\frac{n}{2}\log\left[n(\alpha-1)+2\alpha\right]+\log(\sigma_1\cdots\sigma_n) \quad (32)$$

$$\frac{n}{n+2}<\alpha<1$$

But

$$\operatorname*{Lt}_{\alpha\to1}\ (1/1-\alpha)\log\frac{2\alpha}{n(\alpha-1)+2\alpha}=\frac{n}{2} \quad (33)$$

$$\operatorname*{Lt}_{\alpha\to1}\ \log\left[\beta\left(\frac{1}{1-\alpha}-\frac{1}{2},\frac{1}{2}\right)\cdots\beta\left(\frac{1}{1-\alpha}-\frac{n}{2},\frac{1}{2}\right)\right]$$

$$=\operatorname*{Lt}_{\alpha\to1}\log\beta\left(\frac{1}{\alpha-1}-1/2,\ 1/2\right)+\cdots+\operatorname*{Lt}_{\alpha\to1}\log\beta\left(\frac{1}{1-\alpha}-\frac{n}{2},\frac{1}{2}\right) \quad \ldots (34)$$

By Gauss definition of Gamma function, page 770, Kreyszig (1962), we have

$$\Gamma\alpha=\operatorname*{Lt}_{n\to\infty}\frac{n!\,n^\alpha}{\alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n)} \qquad \alpha>0$$

Using this (34) would reduce to

$$\underset{\alpha \to 1}{\mathrm{Lt}} \log \left[\beta\left(\frac{1}{1-\alpha}-\frac{1}{2}, \frac{1}{2}\right)\ldots\beta\left(\frac{1}{1-\alpha}-\frac{n}{2}, \frac{1}{2}\right)\right]$$

$$= \underset{\alpha \to 1}{\mathrm{Lt}} \log (1-\alpha)^{n/2} + n \log \Gamma\tfrac{1}{2} \qquad (35)$$

Using (33), (34) and (35), in the limit, as $\alpha \to 1$, (32) tends to

$$\max H_1(\tilde{X}) = (n/2)+(n/2)\log 2 + \log(\sigma_1 \cdots \sigma_n) + n \log\sqrt{\pi}$$

$$= \log\sqrt{2\pi e}\,\sigma_1 + \ldots + \log\sqrt{2\pi e}\,\sigma_n$$

as already known under the given constraints on $\tilde{X}$.

4. INVALIDITY

It was pointed out in section 3.1, part II, chapter 5, that

$$H_\alpha(\tilde{X}) \leq H_\alpha(X_1) + \ldots + H_\alpha(X_n) \qquad \alpha > 0, \alpha \neq 1 \qquad (36)$$

is not true in general. The following is its proof.

Proof:

Let us consider the result given by (32). This gives the maximum value of $H_\alpha(\tilde{X})$ when each component of $\tilde{X}$ has mean zero and a specified standard deviation. Had (36) been true in general we would have $H_\alpha(\tilde{X})$ to be maximum when

(i) all the dimensions are independent

(ii) each dimension has the greatest α-entropy under the constraints,

and the result would have been

$$\max H_\alpha(\tilde{X}) = n\Big[(1/1-\alpha)\log(2\alpha/3\alpha-1)+(1/2)\log(3\alpha-1)-(1/2)\log(1-\alpha)+\log\beta\left(\frac{\alpha+1}{2(1-\alpha)},\ 1/2\right)\Big]$$

$$+\log(\sigma_1 \ldots\ldots \sigma_n) \qquad (37)$$

since max. $H_\alpha(X_1)$, by putting n =1 in (32) is

$$\frac{1}{1-\alpha}\log\frac{2\alpha}{3\alpha-1}+\frac{1}{2}\log(3\alpha-1)-\frac{1}{2}\log(1-\alpha)+\log\beta\left(\frac{\alpha+1}{2(1-\alpha)},\frac{1}{2}\right)+\log\sigma_1$$

$$\text{for } 1/3 < \alpha < 1$$

A comparison of (32) and (37) shows that it is not so. Hence we can say (36) is not true in general at least for $(n/n+2) < \alpha < 1$. Hence it is not true in general for all α, $\alpha > 0$.

5. EVALUATIONS

In 5.1 α-informations and in 5.2 α-capacities for specific multidimensional continuous channels are evaluated.

5.1 α-Informations When $\tilde{X}$ and $\tilde{Y}$ are Gaussian

From section 2.4, part II, chapter 5 we have for such a case for $\alpha > 0$, $\alpha \neq 1$,

$$H_\alpha(\tilde{Y}) - H_\alpha(\tilde{Y}/\tilde{X}) = H_\alpha(\tilde{X}) - H_\alpha(\tilde{X}/\tilde{Y})$$

and in the limit as $\alpha \to 1$

$$H_1(\tilde{Y}) - H_1(\tilde{Y}/\tilde{X}) = H_1(\tilde{X}) - H_1(\tilde{X}/\tilde{Y})$$

and this is known to be true. So for $\alpha > 0, \alpha \neq 1$

$$I_\alpha(\tilde{X} ; \tilde{Y}) = I_\alpha(\tilde{Y}; \tilde{X}) = I_1(\tilde{X}, \tilde{Y})$$

and we know from Gel'fand and Laglom (1959) that

$$I_1(\tilde{X}, \tilde{Y}) = \frac{1}{2} \log \frac{\det A \det B}{\det C}$$

where A, B and C are respectively the moment matrices of the distributions of $\tilde{X}$, $\tilde{Y}$ and their join.

5.2 α-Capacities

In (i) expression for α-capacity of type II is derived when the channel is affected by additive Gaussian noise and each component of the input vector is distributed with mean zero and a specified standard deviation and in (ii) α- capacities will be considered when besides conditions in (i), input is also Gaussian.

(i) Let the input be represented by the random vector $\tilde{X} = [X_1, \ldots, X_n]$ output by $\tilde{Y} = [Y_1, \ldots, Y_n]$ and noise by $\tilde{Z} = [Z_1, \ldots, Z_n]$.

The assumptions are :

1. $E(X_k) = 0, \; E(X_k^2) = \sigma^2_{x_k}, \quad k = 1, \ldots, n$

2. $E(Z_k) = 0, \; E(Z_k^2) = \sigma^2_{z_k}, \; E(Z_j Z_k) = 0 \; j \neq k$

and $$\phi(\tilde{z}) = \prod_{k=1}^{n} \left[\frac{1}{\sqrt{2\pi}\, \sigma_{z_k}} \exp. \left(-z_k^2 / 2 \sigma^2_{z_k}\right) \right]$$

3. $\tilde{Y} = \tilde{X} + \tilde{Z}$

By assumptions 2 and 3

$$f(\tilde{y}/\tilde{x}) = f(\tilde{x} + \tilde{z}/\tilde{x}) = \phi(\tilde{z})$$

Using this we have

$$H_\alpha(\tilde{Y}/\tilde{X}) = H_\alpha(\tilde{Z})$$

Hence, using this for $\alpha > 0$, $\alpha \neq 1$,

$$I_\alpha(\tilde{Y};\tilde{X}) = H_\alpha(\tilde{Y}) - H_\alpha(\tilde{Y}/\tilde{X}) = H_\alpha(\tilde{Y}) - H_\alpha(\tilde{Z}) \tag{38}$$

Using assumption 2 we shall have

$$H_\alpha(\tilde{Z}) = \sum_{k=1}^{n} \log(\sqrt{2\pi}\,\sigma_{z_k}) + \frac{n}{2(1-\alpha)} \log \frac{1}{\alpha} \tag{39}$$

Using (2), (38) and (39) we have

$$C_\alpha^{(2)} = \max I_\alpha(\tilde{Y};\tilde{X}) = \max H_\alpha(\tilde{Y}) - \left[\sum_{k=1}^{n} \log\sqrt{2\pi}\,\sigma_{z_k} + \frac{n}{2(1-\alpha)} \log \frac{1}{\alpha}\right] \quad \ldots. \ (40)$$

Since each component of $\tilde{X}$ is affected by an independent Gaussian perturbation i.e.

$$Y_k = X_k + Z_k \ ,$$

so according to this and assumption 1, Y_k is distributed with mean zero and standard deviation $\sqrt{(\sigma^2_{x_k} + \sigma^2_{z_k})}$. So the problem of maximising $H_\alpha(\tilde{Y})$ conforms to the problem discussed

as case (iii) in section 3. Hence using (32), (40) reduces to

$$C_\alpha^{(2)} = \left[\frac{1}{1-\alpha} \log \frac{2\alpha}{n(\alpha-1)+2\alpha} + \log(\beta(\frac{1}{1-\alpha} - \frac{1}{2}, \frac{1}{2})\ldots \right.$$

$$\ldots\beta(\frac{1}{1-\alpha} - \frac{n}{2}, \frac{1}{2})) - \frac{n}{2}\log(1-\alpha) + \frac{n}{2}\log(n(\alpha-1)+2\alpha)$$

$$\left. + \frac{n}{2(1-\alpha)} \log \alpha \right]$$

$$+ \sum_{k=1}^{n} \log\sqrt{(\sigma^2_{x_k} + \sigma^2_{z_k})} - \sum_{k=1}^{n} \log\sqrt{2\pi}\,\sigma_{z_k} \quad (41)$$

$$(n/n+2) < \alpha < 1$$

Using (33), (34) and (35) the expression in the bracket on the right hand side of (41) tends to $(n/2) \log 2 + n \log\sqrt{\pi}$ So in the limit, as $\alpha \to 1$, (41) tends to

$$C_1 = (1/2) \sum_{k=1}^{n} \log (1 + \sigma^2_{x_k} / \sigma^2_{z_k})$$

as already known.

(ii) In this case the steps for getting the result given in (40) are the same as those in case (i) above. Since the input is also Gaussian, so $\tilde{Y}$ is normally distributed with $\sigma^2_{y_k} = \sigma^2_{x_k} + \sigma^2_{z_k}$, $k = 1, \ldots, n$. Hence by (71) of part II, chapter 5, $H_\alpha(\tilde{Y})$ is maximum when (1) all the dimensions of $\tilde{Y}$ are independent random variables and (2) each dimension has the greatest α-entropy; but Y_k's are normally distributed with mean zero and standard deviation $\sqrt{(\sigma^2_{x_k} + \sigma^2_{z_k})}$. So calculations would give

$$H_\alpha(Y_k) = \log\sqrt{2\pi(\sigma^2_{x_k} + \sigma^2_{z_k})} + \frac{1}{2(1-\alpha)}\log\frac{1}{\alpha}$$

and hence

$$\max H_\alpha(\tilde{Y}) = \sum_{k=1}^{n} \log\sqrt{2\pi(\sigma^2_{x_k} + \sigma^2_{z_k})} + \frac{n}{2(1-\alpha)}\log\frac{1}{\alpha} \qquad (42)$$

Using (42), (40) reduces to

$$C_\alpha^{(2)} = \frac{1}{2}\sum_{k=1}^{n} \log\left(1+\frac{\sigma^2_{x_k}}{\sigma^2_{z_k}}\right) \quad \alpha > 0,\ \alpha \neq 1 \quad (43)$$

Similarly we shall have

$$C_\alpha^{(1)} = \frac{1}{2}\sum_{k=1}^{n} \log\left(1+\frac{\sigma^2_{x_k}}{\sigma^2_{z_k}}\right) \quad \alpha > 0,\ \alpha \neq 1 \quad (44)$$

Hence in this case both $C_\alpha^{(1)}$ and $C_\alpha^{(2)}$ turn out to be same as the known value of C_1, corresponding Shannon's capacity.

6. A GENERALISATION OF SHANNON-HARTLEY CHANNEL-CAPACITY FORMULA

If we consider a class of signals represented by time functions f(t) which are band limited in the range $-\omega$ to $+\omega$ cycles per second and if we further assume that these signals are also time limited i.e. f(t) is negligible outside a time interval $(-T/2, T/2)$, T being an integer, then we know that such signals with their power content equal to a preassigned number are represented by points in a $2\omega T$-dimensional space.

Now a point in this signal space may be put as

$$\tilde{X} = \left[X_1, X_2, \dots\dots, X_{2wT}\right]$$

Similarly the output signals may be represented by points belonging to another 2w T- dimensional space. So a point from this output signal space may be put as

$$\tilde{Y} = \left[Y_1, Y_2, \dots\dots, Y_{2wT}\right]$$

So the study of the transmission of band- and time-limited signals tantamounts to the study of the behaviour of a 4w T-dimensional random variable $\tilde{Z} = \left[\tilde{X}, \tilde{Y}\right]$. Proceeding exactly the way adopted for getting (2), part II, chapter 5, and following the manner of section 2.1, Part II, chapter 5, we can define various α-entropies connected with this 4w T-dimensional random variable $\tilde{Z}$. Hence under the conditions subject to which (41) is derived we shall have α-capacity of type II, $(wT)/(wT+1) < \alpha < 1$, for band-and time-limited signals i.e.,

$$C_{\alpha}^{(2)} = \Bigg[\frac{1}{1-\alpha}\log\frac{\alpha}{wT(1-\alpha)+\alpha} + \log\Big(\beta\Big(\frac{1}{1-\alpha}-\frac{1}{2}, \frac{1}{2}\Big)\dots$$

$$\dots\beta\Big(\frac{1}{1-\alpha}-wT, 1/2\Big)\Big)$$

$$- wT\log(1-\alpha) + wT\log\left(2wT(\alpha-1)+2\alpha\right) + \frac{wT}{1-\alpha}\log\alpha\Bigg]$$

$$+ \sum_{k=1}^{2wT}\log\sqrt{\sigma^2_{x_k} + \sigma^2_{z_k}} - \sum_{k=1}^{2wT}\log\sqrt{2\pi}\,\sigma_{z_k} \qquad (45)$$

$$(wT/wT+1) < \alpha < 1 .$$

In particular when

$$\sigma^2_{x_k} = \sigma^2_x \qquad k = 1, \ldots, n$$
$$\sigma^2_{z_k} = \sigma^2_z$$

(45) reduces to

$$C_\alpha^{(2)} = \Big[\frac{1}{1-\alpha} \log \frac{\alpha}{wT(1-\alpha) + \alpha} + \log(\beta(\frac{1}{1-\alpha} - \frac{1}{2}, \frac{1}{2}) \ldots$$
$$\ldots \beta(\frac{1}{1-\alpha} - wT, \frac{1}{2}))$$
$$- wT\log(1-\alpha) + wT \log(2wT(1-\alpha) + 2\alpha) + \frac{wT}{1-\alpha} \log \alpha \Big]$$
$$+ wT\log(\sigma^2_x + \sigma^2_z) - wT \log(2\pi \sigma^2_z) \tag{46}$$
$$\frac{wT}{wT+1} < \alpha < 1$$

Using (33), (34) and (35) it can be easily verified that in the limit, as $\alpha \to 1$, (46) tends to a known form of Shannon-Hartley formula.

Under the conditions for which case (ii) in section 5.2 has been tackled it can be easily verified that both $C_\alpha^{(1)}$ and $C_\alpha^{(2)}$ would turn out to be same as a known form of Shannon- Hartley formula under these conditions.

7. ALTERNATIVE DEFINITION OF CONDITIONAL α-ENTROPY

As developed in section 5, part II, chapter 5, there are alternative definitions of conditional α-entropy and α-equivocation for multidimensional continuous channels. These are respectively given by (81), (82)-part II, chapter 5.

It may be noted that all the results, that involve these concepts of conditional α-entropy and α-equivocation and are developed in this chapter, are also true for these alternative definitions. This can be easily verified.

CHAPTER - 7

SEMI-CONTINUOUS MULTIDIMENSIONAL MEMORYLESS CHANNELS

1. INTRODUCTION

As defined in Feinstein (1958) a semi-continuous memoryless channel is a channel with input consisting of a finite set and output an infinite set with a probability measure defined on the output for every input, so a multi-dimensional semi-continuous memoryless channel is that for which the input consists of a finite set of n-tuples and output an infinite set of m-tuples with the further restriction that each transmitted n-tuple is independent of the previously transmitted n-tuple i.e. no memory is involved; however components of a particular n-tuple may be interdependent. Let the input consist of s n-tuples $\breve{x}^{(1)}, \ldots, \breve{x}^{(s)}$ and let $\breve{X} = [X_1, \ldots, X_n]$ be a random variable which takes on the value $\breve{x}^{(i)}$ with probability $p(\breve{x}^{(i)}) > 0$ so that $\sum_{i=1}^{s} p(\breve{x}^{(i)}) = 1$. Let the output be described by $\breve{Y} = [Y_1, \ldots, Y_m]$ having the density function $p(\breve{y})$, positive and defined over an m-dimensional rectangle R_m so that $\int_{R_m} p(\breve{y})\, d\breve{y} = 1$. The case when $p(\breve{y})$ vanishes at the most over a set of measure zero can be included either by adopting Shannon's convension that $0 \times \infty = 0$ or by evaluating the various α-entropies associated with such a channel over the extended real number system. As for a multidimensional continuous channel, there are also five α-entropies associated with a channel under consideration.

(i) Associated with the input

$$H_\alpha(\tilde{X}) = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^{s} p^\alpha(\tilde{x}^{(i)}) \qquad \alpha > 0, \alpha \neq 1 \quad (1)$$

(ii) Associated with the output

$$H_\alpha(\tilde{Y}) = \frac{1}{1-\alpha} \log_2 \int_{R_m} p^\alpha(\tilde{y})\, d\tilde{y} \qquad \alpha > 0, \alpha \neq 1 \quad (2)$$

(iii) Associated with the system as a whole

$$H_\alpha(\tilde{X}, \tilde{Y}) = \frac{1}{1-\alpha} \log_2 \sum_i \int_{R_m} p^\alpha(\tilde{x}^{(i)}, \tilde{y})\, d\tilde{y} \qquad \alpha > 0, \alpha \neq 1 \quad (3)$$

(iv) Conditional α-entropy $H_\alpha(\tilde{Y}/\tilde{X})$. Since a transmitted vector $\tilde{x}^{(i)}$ can be received as any admissible value of $\tilde{Y}$ and $\int_{R_m} f(\tilde{y}/\tilde{x}^{(i)})\, d\tilde{y} = 1$, so following the procedure of getting (2) we shall have

$$H_\alpha(\tilde{Y}/\tilde{x}^{(i)}) = \frac{1}{1-\alpha} \log_2 \int_{R_m} f^\alpha(\tilde{y}/\tilde{x}^{(i)})\, d\tilde{y}$$

So for $H_\alpha(\tilde{Y}/\tilde{X})$, it is, as in earlier chapters, appropriate to take the mathematical expectation $H_\alpha(\tilde{Y}/\tilde{x}^{(i)})$. Hence

$$H_\alpha(\tilde{Y}/\tilde{X}) = \frac{1}{1-\alpha} \sum_i p(\tilde{x}^{(i)}) \log_2 \int_{R_m} f^\alpha(\tilde{y}/\tilde{x}^{(i)})\, d\tilde{y} \quad (4)$$

$\alpha > 0, \alpha \neq 1$

(v) α-equivocation $H_\alpha(\tilde{X}/\tilde{Y})$. Based on the procedure adopted for getting (4) we shall have

$$H_\alpha(\tilde{X}/\tilde{Y}) = \frac{1}{1-\alpha} \int_{R_m} f(\tilde{y}) \left(\log_2 \sum_i p^\alpha(\tilde{x}^{(i)}/\tilde{y}) \right) d\tilde{y} \quad (5)$$

$\alpha > 0, \alpha \neq 1$

It can be easily verified that in the limit, as $\alpha \to 1$, (1), (2), (3), (4) and (5) tend to the known entropies for such channels.

All the other results proved in sections 2 and 3, part II, chapter 5 for multidimensional continuous memoryless channels can similarly be proved for such channels. The concepts of α-informations and α-capacities defined for multidimensional continuous channels can likewise be defined for such channels.

In section 2 available Fano's inequality for such channels is generalised. As a consequence of this generalisation, the converse of Shannon's second fundamental theorem for semi-continuous memoryless channels is generalised in section 3. In section 4 these results are considered in the light of alternative definitions of conditional α-entropy and α-equivocation.

The operations dealt with below are contingent upon the existence of underlying integrals and the application of ''Jensen's inequality for multiple integrals'' is made subject to the restrictions as stated for its validity in part II, chapter 5. For the rest the base of the logs is assumed to be e.

2. GENERALISATION OF FANO'S INEQUALITY

We know that for such channels Fano's inequality is given by

$$H_1(\tilde{X}/\tilde{Y}) \leq H_1(p(e),\ 1-p(e)) + p(e) \log (s-1) \qquad (6)$$

where p(e), the probability of error of the code consisting of s words $\tilde{x}^{(1)}, \ldots\ldots \tilde{x}^{(s)}$, is given by

$$p(e) = \int_{R_m} p(\tilde{y}) \left[1 - p\,(g(\tilde{y})/\ \tilde{Y} = \tilde{y})\right] d\tilde{y}$$

$$= 1 - \int_{R_m} p(g(\tilde{y}),\ \tilde{y})\ d\tilde{y} \qquad (7)$$

(7) is calculated on the assumption that the decoding scheme is such that assigns the vector $g(\tilde{y}) = \tilde{x}^*$(say) to a received vector $\tilde{y}$. With this (7) can be written as

$$p(e) = \sum_{\{\tilde{X}\} - \tilde{x}^*} \int_{R_m} f(\tilde{x},\ \tilde{y})\ d\tilde{y} \qquad (8)$$

where $\{\tilde{X}\}$ stands for the set of s code words. From (8) we have

$$1 - p(e) = \int_{R_m} p(\tilde{x}^*,\ \tilde{y})\ d\tilde{y} \qquad (9)$$

For α-entropies (6) will get generalised to

$$H_\alpha(\tilde{X}/\tilde{Y}) \leqslant (1/1-\alpha)\log\left[p^\alpha(e)(s-1)^{1-\alpha} + (1-p(e))^\alpha\right] \qquad (10)$$

$$\alpha > 0,\ \alpha \neq 1.$$

It can be easily verified that in the limit, as $\alpha \to 1$, (10) tends to (6).

<u>Proof of (10)</u> :

<u>Case When $0 < \alpha < 1$</u>

$H_\alpha(\tilde{X}/\tilde{Y})$ given by (5) can be written as

$$H_\alpha(\tilde{X}/\tilde{Y}) = (1/1-\alpha) \int_{R_m} p(\tilde{y}) \left[\log \sum_{\{\tilde{X}\}} p^\alpha(\tilde{x}\ /\tilde{y})\right] d\tilde{y} \qquad (11)$$

But by Jensey's inequality for multiple integrals, page 21, Morrey (1966) we have

$$\int_{R_m} p(\tilde{y}) \left[\log \sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y})\right] d\tilde{y} \leq \log \int_{R_m} p(\tilde{y}) \left[\sum_{\{\tilde{X}\}-\tilde{x}^*} p^{\alpha}(\tilde{x}/\tilde{y}) + p^{\alpha}(\tilde{x}^*/\tilde{y})\right] d\tilde{y}$$

$$= \log \left[\int_{R_m} \sum_{\{\tilde{X}\}-\tilde{x}^*} p(\tilde{y}) p^{\alpha}(\tilde{x}/\tilde{y}) d\tilde{y} + \int_{R_m} p(\tilde{y})\, p^{\alpha}(\tilde{x}^*/\tilde{y})\, d\tilde{y}\right]$$

$$= \log \left[\int_{R_m} \sum_{\{\tilde{X}\}-\tilde{x}^*} p^{\alpha}(\tilde{y}) p^{\alpha}(\tilde{x}/\tilde{y})\, p^{1-\alpha}(\tilde{y})\, d\tilde{y} + \int_{R_m} p^{\alpha}(\tilde{y})\, p^{\alpha}(\tilde{x}^*/\tilde{y})\, p^{1-\alpha}(\tilde{y})\, d\tilde{y}\right]$$

$$= \log \left[\int_{R_m} \sum_{\{\tilde{X}\}-\tilde{x}^*} p^{\alpha}(\tilde{x}, \tilde{y})\, p^{1-\alpha}(\tilde{y})\, d\tilde{y} + \int_{R_m} p^{\alpha}(\tilde{x}^*, \tilde{y})\, p^{1-\alpha}(\tilde{y})\, d\tilde{y}\right] \tag{12}$$

But by Hölder inequality for sums, page 21, Beckenbach and Bellman (1965),

$$\sum_{\{\tilde{X}\}-\tilde{x}^*} p^{\alpha}(\tilde{x},\tilde{y}) p^{1-\alpha}(\tilde{y}) \leq \Big(\sum_{\{\tilde{X}\}-\tilde{x}^*} p(\tilde{x},\tilde{y})\Big)^{\alpha} \big((s-1)\, p(\tilde{y})\big)^{1-\alpha} \tag{13}$$

Integrating (13) over R_m we have

$$\int_{R_m} \Big(\sum_{\{\tilde{X}\}-\tilde{x}^*} p^{\alpha}(\tilde{x},\tilde{y}) p^{1-\alpha}(\tilde{y})\Big) d\tilde{y} \leq \int_{R_m} \Big(\sum_{\{\tilde{X}\}-\tilde{x}^*} p(\tilde{x}, \tilde{y})\Big)^{\alpha} \big((s-1) p(\tilde{y})\big)^{1-\alpha} d\tilde{y} \quad \ldots (14)$$

Applying Holder inequality for multiple integrals Beckenbach and Bellman (1965) to the R.H.S. of (14), it reduces to

$$\int_{R_m} \Big(\sum_{\{\tilde{X}\} - \tilde{x}^*} p^{\alpha}(\tilde{x}, \tilde{y})\, p^{1-\alpha}(\tilde{y}) \Big)\, d\tilde{y}$$

$$\leq \Big(\int_{R_m} \sum_{\{\tilde{X}\} - \tilde{x}^*} p(\tilde{x},\tilde{y})\,d\tilde{y} \Big)^{\alpha} \Big(\int_{R_m} (s-1)\, p(\tilde{y})\, d\tilde{y} \Big)^{1-\alpha} \qquad (15)$$

Again by Hölder inequality for multiple integrals

$$\int_{R_m} p^{\alpha}(\tilde{x}^*, \tilde{y})\, p^{1-\alpha}(\tilde{y})\,d\tilde{y} \leq \Big(\int_{R_m} p(\tilde{x}^*, \tilde{y})\, d\tilde{y} \Big)^{\alpha} \qquad (16)$$

Using (15) and (16), (12) reduces to

$$\int_{R_m} p(\tilde{y}) \Big[\log \sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y}) \Big]\, d\tilde{y}$$

$$\leq \log \Big[\Big(\int_{R_m} \sum_{\{\tilde{X}\} - \tilde{x}^*} p(\tilde{x},\tilde{y})\, d\tilde{y} \Big)^{\alpha} (s-1)^{1-\alpha} + \Big(\int_{R_m} p(\tilde{x}^*,\tilde{y})\,d\tilde{y} \Big)^{\alpha} \Big] \qquad (17)$$

Multiplying (17) by $(1/1-\alpha)$, a positive number, and using (8), (9) and (11) we shall have (10) for $0 < \alpha < 1$.

<u>Case When $\alpha > 1$</u>

Since $\sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y}) \leq 1$, so by Jensen's inequality for multiple integrals

$$\int_{R_m} p(\tilde{y})(\log \sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y}))d\tilde{y} \geqslant \int_{R_m} p(\tilde{y})(\sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y})\log \sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y}))d\tilde{y}$$

$$\geqslant \left[\int_{R_m} p(\tilde{y})(\sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y}))d\tilde{y}\right] \log \left[\int_{R_m} p(\tilde{y})(\sum_{\tilde{X}} p^{\alpha}(\tilde{x}/\tilde{y}))d\tilde{y}\right] \quad \ldots (18)$$

But $\int_{R_m} p(\tilde{y}) (\sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y})) d\tilde{y} \leq 1$, let it be β ;

Using this (18) reduces to

$$\int_{R_m} p(\tilde{y})(\log \sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y}))d\tilde{y} \geqslant \beta \log \left[\int_{R_m} p(\tilde{y})(\sum_{\{\tilde{X}\}} p^{\alpha}(\tilde{x}/\tilde{y}))d\tilde{y}\right]$$

$$= \beta \log \left[\int_{R_m} \sum_{\{\tilde{X}\}-\tilde{x}^*} p(\tilde{y})\, p^{\alpha}(\tilde{x}/\tilde{y})\, d\tilde{y} + \int_{R_m} p(\tilde{y})\, p^{\alpha}(\tilde{x}^*/\tilde{y})\, d\tilde{y}\right]$$

$$= \beta \log \left[\int_{R_m} (\sum_{\{\tilde{X}\}-\tilde{x}^*} p^{\alpha}(\tilde{x},\tilde{y})\, p^{1-\alpha}(\tilde{y}))d\tilde{y} + \int_{R_m} p^{\alpha}(\tilde{x}^*, \tilde{y})\, p^{1-\alpha}(\tilde{y})\, d\tilde{y}\right] \quad (19)$$

But by Hölder inequality for sums

$$\sum_{\{\tilde{X}\}-\tilde{x}^*} p^{\alpha}(\tilde{x},\tilde{y})p^{1-\alpha}(\tilde{y}) \geqslant (\sum_{\{\tilde{X}\}-\tilde{x}^*} p(\tilde{x},\tilde{y}))^{\alpha}((s-1)p(\tilde{y}))^{1-\alpha} \quad (20)$$

Integrating (20) over R_m and then applying Hölder inequality

for multiple integrals to the R.H.S. of the result so obtained we shall have

$$\int_{R_m} \Big(\sum_{\{\bar{X}\}-\bar{x}^*} p^{\alpha}(\bar{x},\bar{y})\, p^{1-\alpha}(\bar{y})\Big) d\bar{y} \geq \Big(\int_{R_m} \sum_{\{\bar{X}\}-\bar{x}^*} p(\bar{x},\bar{y})\, d\bar{y}\Big)^{\alpha}$$

$$\Big(\int_{R_m} (s-1)\, p(\bar{y})\, d\bar{y}\Big)^{1-\alpha}$$

$$= \Big(\int_{R_m} \sum_{\{\bar{X}\}-\bar{x}^*} p(\bar{x}, \bar{y})\, d\bar{y}\Big)^{\alpha} (s-1)^{1-\alpha} \qquad (21)$$

Again by Hölder inequality for multiple integrals

$$\int_{R_m} p^{\alpha}(\bar{x}^*, \bar{y})\, p^{1-\alpha}(\bar{y})\, d\bar{y} \geq \Big(\int_{R_m} p(\bar{x}^*, \bar{y})\, d\bar{y}\Big)^{\alpha} \qquad (22)$$

Using (21) and (22), (19) reduces to

$$\int_{R_m} p(\bar{y}) \Big(\log \sum_{\{\bar{X}\}} p^{\alpha}(\bar{x}/\bar{y})\Big) d\bar{y} \geq \beta \log \Bigg[\Big\{ \int_{R_m} \sum_{\{\bar{X}\}-\bar{x}^*} p(\bar{x},\bar{y})\, d\bar{y} \Big\}^{\alpha} (s-1)^{1-\alpha}$$

$$+ \Big\{ \int_{R_m} p(\bar{x}^*,\bar{y})\, d\bar{y} \Big\}^{\alpha} \Bigg] \qquad (23)$$

Multiplying (23), by $(1/1-\alpha)$, a negative number and using (8), (9) and (11) and the fact that $\beta \leq 1$ we shall have (10) for $\alpha > 1$.

This completes the proof.

3. A GENERALISATION OF THE CONVERSE OF SHANNON'S SECOND FUNDAMENTAL THEOREM FOR SEMI-CONTINUOUS MEMORYLESS CHANNELS

A generalised form of the converse of Shannon's second fundamental theorem is stated and the proof is given for a family of channels consisting of the extensions of any semi-continuous memoryless channel subject to the further restrictions that the input probability distribution is a product distribution and its code words are selected with equal probability. For this we need the following lemma.

Lemma : α-capacity of either type of an nth order extension of a semi-continuous memoryless channel with input probability distribution as a product distribution is n times the α-capacity of the corresponding type of the channel.

The proof for this lemma can be constructed in the same way as that of the corresponding lemma for discrete memoryless channels, section 3.2, chapter 4. Q. E. D.

Below the symbol (s, n) stands for a code with s input n-sequences.

Statement of The Generalised Converse

Given $\epsilon > 0$, it is impossible to transmit a code (s,n) with $s = 2^{n(C_\alpha^{(1)} + \epsilon)}$ through an nth extension of a semi-continuous memoryless noisy channel having α-capacity $C_\alpha^{(1)}$ with as small a probability of error as

desired.

As $\alpha \to 1$, the above statement tends to an already available statement.

Using (10) and the above lemma and following the lines of proof of the corresponding generalisation given for discrete memoryless channels, section 3.2, chapter 4, the proof of the above statement, with the additional assumptions that the input probability distribution of the nth extension is a product distribution and all the s input n-sequences are equally likely, can easily be constructed.

4. ALTERNATIVE DEFINITIONS OF CONDITIONAL α-ENTROPY AND α-EQUIVOCATION

Following the procedures adopted for getting (81) and (82), part II, chapter 5, the alternative definitions for conditional α-entropy and α-equivocation for channels under consideration are given by

$$H_\alpha(\tilde{Y}/\tilde{X}) = (1/1-\alpha)\log \sum_i p(\tilde{x}^{(i)}) \left[\int_{R_m} p^\alpha(\tilde{y}/\tilde{x}^{(i)})d\tilde{y}\right] \quad \alpha > 0, \alpha \neq 1 \tag{24}$$

$$H_\alpha(\tilde{X}/\tilde{Y}) = (1/1-\alpha)\log \int_{R_m} p(\tilde{y})(\sum_i p^\alpha(\tilde{x}^{(i)}/\tilde{y}))\, d\tilde{y} \quad \alpha > 0, \alpha \neq 1$$

It may be noted that all the results developed in this chapter are also true for these alternative definitions. For instance, the truth of (10) for $H_\alpha(\tilde{X}/\tilde{Y})$ given by (24) can be seen as follows :

Case When $0 < \alpha < 1$

As in (12) we have

$$\log \int_{R_n} p(\bar{y}) \left[\sum_{\{\bar{X}\}} p^{\alpha}(\bar{x}/\bar{y}) \right] d\bar{y} = \log \left[\int_{R_n} \sum_{\{\bar{X}\} - \bar{x}^*} p^{\alpha}(\bar{x},\bar{y})\, p^{1-\alpha}(\bar{y})\, d\bar{y} \right.$$

$$\left. + \int_{R_n} p^{\alpha}(\bar{x}^*, \bar{y})\, p^{1-\alpha}(\bar{y})\, d\bar{y} \right]$$

Using (15) and (16), this reduces to

$$\log \int_{R_n} p(\bar{y}) \left[\sum_{\{\bar{X}\}} p^{\alpha}(\bar{x}/\bar{y}) \right] d\bar{y} \leq \log \left[\left(\int_{R_n} \sum_{\{\bar{X}\} - \bar{x}^*} p(\bar{x}_*,\bar{y})\, d\bar{y} \right)^{\alpha} (s-1)^{1-\alpha} \right.$$

$$\left. + \left(\int_{R_n} p(\bar{x}^*, \bar{y})\, d\bar{y} \right)^{\alpha} \right] \qquad (25)$$

Multiplying (25) by $(1/1-\alpha)$, a positive number, and using (8), (9) and (11) we shall have (10) for this case.

Case When $\alpha > 1$

Using (21) and (22), (24) reduces to

$$\log \int_{R_n} p(\bar{y}) \left[\sum_{\{\bar{X}\}} p^{\alpha}(\bar{x}/\bar{y}) \right] d\bar{y} \geq \log \left[\left(\int_{R_n} \sum_{\{\bar{X}\} - \bar{x}^*} p(_*\bar{x},\bar{y})\, d\bar{y} \right)^{\alpha} (s-1)^{1-\alpha} \right.$$

$$\left. + \left(\int_{R_n} p(\bar{x}^*, \bar{y})\, d\bar{y} \right)^{\alpha} \right] \qquad (26)$$

Multiplying (26) by $(1/1-\alpha)$, a negative number, and using (8), (9) and (11) we shall have (10) for this case.

CHAPTER 8

STATIONARY MARKOV CHAINS AND STATIONARY SYMMETRIC CHANNELS

INTRODUCTION

This chapter is also divided into two parts. Part I deals with the definitions of α-entropies associated with different steps of a Morkov chain and some relations between them. These relations tend in the limit as $\alpha \to 1$ to the available relation connecting Shannon's entropies associated with different steps. In part II α-entropies associated with a stationary source and basic concepts like rates of transmission of α-informations through a stationary symmetric channel are developed.

PART I

1. STATIONARITY, NOTATIONS AND α-ENTROPIES

In section 1.1 concept of stationarity of a Markov chain is defined and some notations are given. In section 1.2 α-entropies associated with different steps are given.

1.1 Stationarity And Notations

Suppose a sequence of trials is performed and the result of each is n mutually exclusive events; let, in particular, the sth trial result in n mutually exclusive events $A_1^{(s)},\ldots\ldots, A_n^{(s)}$. A homogeneous or stationary Markov chain is one in which the conditional probability of occurrence of the event $A_i^{(s+1)}$ in the (s+1)th trial, given that the event $A_i^{(s)}$ has been realized in the sth trial, is

independent of the number of the trial ; the probability is called the probability of transition and is usually denoted by p_{ij} i.e. the probability of going from the ith event to the jth event.

Below these events will be named as states and p_{ij} will then be interpreted as the transition probability of going from the ith state to the jth state. Since our dealings are to be confined to stationary Markov chains, so it is sufficient to denote the mutually exclusive events associated with it, called states from onward, without the superscript (s) i.e., by simply $A_1, A_2, \ldots, A_n$.

1.2 <u>α-Entropies</u>

Let $A_1, \ldots, A_n$ be the finite number n of states of a chain and let the matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

where $p_{ij} > 0$

$$\text{and } \sum_{j=1}^{n} p_{ij} = 1 , \; i = 1,2,\ldots.n \tag{1}$$

be its transition probability matrix

By (1) , the subset of transition probabilities from the state A_1 form a complete distribution, so we can define one-step α-entropy, to be denoted by $H_{1\alpha}^{(1)}$,

$$H_{i\alpha}^{(1)} = (1/1-\alpha) \log \sum_{j=1}^{n} p_{ij}^{\alpha} \quad , \alpha > 0, \alpha \neq 1 .$$

If the probability of the system being initially in state A_i is designated by p_i i.e. if $[p_1,\ldots\ldots,p_n]$ is a specified set of initial state probabilities, so α-entropy or α-uncertainty of the chain should naturally be defined as the average of $H_{i\alpha}^{(1)}$ i.e. the α-entropy of moving one step ahead from any initial state. So denoting it by $H_{\alpha}^{(1)}$, we have

$$H_{\alpha}^{(1)} = \sum_{i=1}^{n} p_i \left(1/1-\alpha \;\; \log \sum_{j=1} p_{ij}^{\alpha} \right)$$

or

$$H_{\alpha}^{(1)} = (1/1-\alpha) \sum_{i=1}^{n} p_i \log \sum_{j=1}^{n} p_{ij}^{\alpha} , \qquad (2)$$

$$\alpha > 0, \alpha \neq 1.$$

We may call $H_{\alpha}^{(1)}$, the α-entropy of the chain with specified initial probabilities for moving one step,

In general, the set of events of going from A_1 to any other state in r steps i.e.,

$$\left[A_1^{(r)}/A_1 \, , A_2^{(r)}/A_1, \ldots\ldots\ldots\ldots, A_n^{(r)}/A_1 \right]$$

constitutes a finite probability scheme with n^r events. The r-step α-entropy of this finite scheme is, denoting it by $H_{\alpha 1}^{(r)}$,

$$H_{i\alpha}^{(r)} = (1/1-\alpha) \log \sum_j (p_{ij}^{r)})^{\alpha} \qquad \alpha > 0, \alpha \neq 1$$

where $p_{ij}^{(r)} > 0$ is the probability of moving from ith to the jth state in r-steps. Hence the r-step α-entropy is

$$H_{\alpha}^{(r)} = \sum_{i=1}^{n} p_i H_{i\alpha}^{(r)}$$

$$H_{\alpha}^{(r)} = (1/1-\alpha) \sum_{i=1}^{n} p_i \log \sum_{j=1}^{n} (p_{ij}^{(r)})^{\alpha} \tag{3}$$

$$\alpha > 0, \alpha \neq 1$$

where $[p_1, \ldots\ldots p_n]$ is the specified set of initial probabilities.

In the limit as $\alpha \to 1$, (2) and (3) tend to the corresponding Shannon's definitions.

2. RELATIONS IN α-ENTROPIES

Some relations between α-entropies associated with different steps are given in the form of two theorems since the results of theorem 1 are true only for regular Markov chains, so we shall enunciate these terms in section 2.1 and the theorems would be given in section 2.2.

2.1 Some Known Results

(1) We define a chain with transition probability matrix P a regular chain if there exists a positive integer r such that P^r has only strictly positive elements and

(2) the steady state probabilities arise if for some

n, may be very large, P^n has almost identical rows i.e., the prob. of being in any state s_j (say) at any time t= n (or after n-steps) is almost independent of the starting state i.e. at t = 0. Mathematically we can express this condition by saying that

$$\underset{n \to \infty}{\mathrm{Lt}}\ p_{ij}^{(n)} = w_j \qquad (i, j = 1,2,\ldots\ldots, n)$$

where w_j is independent of i. The number w_j is called the steady state probability of the state s_j. So in a few words if there is a set of numbers $w_1, \ldots\ldots, w_n$ such that $\underset{n \to \infty}{\mathrm{Lt}}\ p_{ij}^{(n)} = w_j$ $(i,j = 1,2,\ldots\ldots, n)$ then we say that steady state probabilities exist for the chain.

(3) Another result that we require is the following theorem, page 174, Ash (1966).

''If a Markov chain with transition matrix P has steady state probabilities w_j $(j =1,2,\ldots\ldots,n)$ then (i) $\sum_{j=1}^{n} w_j = 1$ (ii) $W = [w_1, \ldots\ldots, w_n]$ is a ''stationary distribution'' for the chain i.e. $WP = W$ (iii) this stationary distribution W is unique .

2.2 Theorems

Th.1 If a chain has all $p_{ij} > 0$ in its transition probability matrix and $[p_1, \ldots\ldots, p_n]$ are its steady state probabilities then for $0 < \alpha < 1$.

$$\text{(i)} \quad H_\alpha^{(r)} \geqslant r\, H^{(1)}_{(\alpha - 1/k)k'}$$

where $0 < k < 1$ and k' is so chosen that $(1/k) + (1/k') = 1$

(ii) $H_\alpha^{(r)} \leqslant r \max_i H_{i\alpha}^{(1)}$

and for $\alpha > 1$

(iii) $H_\alpha^{(r)} \leqslant r\ H^{(1)}_{(\alpha-1/k)k'}$

where k and k' are chosen as above

(iv) $H_\alpha^{(r)} \geqslant r \min_i H_{i\alpha}^{(1)}$

<u>Note:</u> that the results (i) and (iii) are infinite sets of inequalities each member of either tending in the limit, as $\alpha \to 1$, to the known equality $H_1^{(r)} = r\ H_1^{(1)}$. As $\alpha \to 1$, results (ii) and (iv) tend respectively to $H_1^{(r)} \leqslant r \max_i H_{i1}^{(1)}$ and $H_1^{(r)} \geqslant r \min_i H_{i1}^{(1)}$.

<u>Proof</u>

Since $p_{ij} > 0$, so the chain is regular and it is given that p_i are steady state probabilities, hence by result (3) stated in section 2.1, we have

$$p_i = \sum_{j=1}^{n} p_j\ p_{ji} \qquad i = 1,\ldots,n \quad (4)$$

(i) the result is trivially true for $r = 1$ i.e.

$$H_\alpha^{(1)} \geqslant H^{(1)}_{(\alpha-1/k)k'} \qquad (5)$$

the equality will occur when p_{ij} are equal. (5) lays down the basis for induction. Hence suppose the result is true for some $r \geqslant 1$ i.e.

$$H_{\alpha}^{(r)} \geqslant r\, H_{(\alpha-1/k)k'}^{(1)} \tag{6}$$

We shall prove that it is true for r+1 i.e.

$$H_{\alpha}^{r+1} \geqslant (r+1)\; H_{(\alpha-1/k)k'}^{(1)} \tag{7}$$

Let the system be in the state A_i. The finite scheme which describes the path of the system in the next (r+1) steps can be regarded as the product of two dependent schemes.

X : the scheme corresponding to the immediately following step with the α-entropy $H_{i\alpha}^{(1)}$ and

Y : the scheme describing the system in the next r steps; the α-entropy of the scheme is $H_{j\alpha}^{(r)}$ if the outcome of the scheme X was the event x_j.

So using (62) of chapter 2 we shall have

$$H_{i\alpha}^{(r+1)} \geqslant H_{i(\alpha-1/k)k'} + \sum_j p_{ij}\, H_{j\alpha}^{(r)} \tag{8}$$

Multiplying 8) by p_i, summing over i we shall have

$$H_{\alpha}^{(r+1)} \geqslant H_{(\alpha-1/k)k'}^{(1)} + \sum_j H_{j\alpha}^{(r)} \left(\sum_i p_i\, p_{ij} \right) \tag{9}$$

Using (4) and (6), (9) reduces to the required result given by (7).

(ii) The result is trivially true for r = 1 i.e.

$$H_{\alpha}^{(1)} \leq \max_{i} H_{i\alpha}^{(1)}$$

Suppose it is true for r = m i.e.

$$H_{\alpha}^{(m)} \leq m \max_{i} H_{i\alpha}^{(1)} \tag{10}$$

If we argue as in result (i), take X to be the scheme describing m steps following A_i, Y describing the m+1 th step following the result of X and x_j be the outcome of X then by (69) of chapter 2 we have

$$H_{i\alpha}^{(m+1)} \leq H_{i\alpha}^{(m)} + \max_{j} H_{j\alpha}^{(1)} \tag{11}$$

Multiplying (11) by p_i , summing over i, changing the suffix j to i and using (10) we shall have

$$H_{\alpha}^{(m+1)} \leq (m+1) \max_{i} H_{i\alpha}^{(1)}$$

whence the required result follows:

Similarly following the above procedure and using (70) and (74) of chapter 2 we can prove results (iii) and (iv).

Q.E.D.

__Th. 2__ If the transition probability matrix $[p_{ij}]$ i,j = 1,.. ...,n of a Markov chain is in the form of a channel matrix with symmetric noise structure then the different steps α- entropies of the chain are additive i.e.

$$H_\alpha^{(r)} = r\,H_\alpha^{(1)} \qquad \alpha > 0, \tag{12}$$

Proof   When $[p_{ij}]$ is of order 1, $H_\alpha^{(r)} = 0$ for all r

When $[p_{ij}]$ is of order 2,

Let

$$P = \begin{array}{c|cc} & A_1 & A_2 \\ \hline A_1 & p_1 & p_2 \\ A_2 & p_2 & p_1 \end{array} \qquad \text{with } p_1 + p_2 = 1$$

In One step a state A_1 can become

$A_1 \to A_1$ with prob. p_1

$A_1 \to A_2$ with prob. p_2

and thus proceeding as in (1)

$$H_\alpha^{(1)} = (1/1-\alpha)\log\left(p_1^\alpha + p_2^\alpha\right) \qquad \alpha > 0,\ \alpha \neq 1$$

In two steps a state A_1 can become

$A_1 \to A_1 \to A_1$ with prob. p_1^2

$A_1 \to A_1 \to A_2$ with prob. $p_1\,p_2$

$A_1 \to A_2 \to A_2$ With prob. $p_2\,p_1$

$A_1 \to A_2 \to A_1$ With prob. p_2^2

But these probabilities are the various terms of $(p_1+p_2)^2$, hence

$$H_\alpha^{(2)} = (1/1-\alpha)\log\left(p_1^\alpha + p_2^\alpha\right)^2 = 2\,H_\alpha^{(1)} \quad \alpha > 0, \alpha \neq 1$$

Similarly in r steps the state A_1 can become

$A_1 \to A_1 \to - - - - - \to A_1$ with prob. p_1^r

$A_1 \to A_1 \to - - - - - \to A_1 \to A_2$ with prob. $p_1^{r-1}\ p_2$

This second event can be had in $\binom{r}{1}$ ways, each with prob. $p_1^{r-1}\ p_2$, since A_1 can go to A_1 in any $r-1$ steps out of the total number of r steps to be taken by it. Similarly probabilities of other events can be had. All these probabilities are nothing but the various terms in $(p_1 + p_2)^r$. Hence

$$H_\alpha^{(r)} = (1/1-\alpha) \log (p_1^\alpha + p_2^\alpha)^r = r\, H_\alpha^{(1)} \qquad \alpha > 0, \alpha \neq 1, \infty$$

<u>When</u> $[p_{ij}]$ <u>is of order 3</u>

$$\text{Let} \qquad P = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2 & p_3 & p_1 \\ p_3 & p_1 & p_2 \end{bmatrix} \end{array} \qquad \text{with } p_1 + p_2 + p_3 = 1$$

$$H_\alpha^{(1)} = (1/1-\alpha) \log (p_1^\alpha + p_2^\alpha + p_3^\alpha)$$

In two steps the state A_1 can become

$A_1 \to A_1 \to A_1$ with prob. p_1^2

$A_1 \to A_1 \to A_2$ with prob. $p_1 p_2$

$A_1 \to A_1 \to A_3$ with prob. $p_1 p_3$

$$A_1 \rightarrow A_2 \rightarrow A_1 \qquad \text{with prob.} \quad p_2^2$$

$$A_1 \rightarrow A_2 \rightarrow A_2 \qquad \text{with prob.} \quad p_2\, p_3$$

$$A_1 \rightarrow A_2 \rightarrow A_3 \qquad \text{with prob.} \quad p_2\, p_1$$

$$A_1 \rightarrow A_3 \rightarrow A_1 \qquad \text{with prob.} \quad p_3^2$$

$$A_1 \rightarrow A_3 \rightarrow A_2 \qquad \text{with prob.} \quad p_3\, p_1$$

$$A_1 \rightarrow A_3 \rightarrow A_3 \qquad \text{with prob.} \quad p_3\, p_2$$

Since these probabilities are the various terms of $(p_1+p_2+p_3)^2$, so

$$H_\alpha^{(2)} = (1/1-\alpha) \log (p_1^\alpha + p_2^\alpha + p_3^\alpha)^2 = 2\, H_\alpha^{(1)} \quad \alpha > 0, \ \alpha \neq 1$$

Similarly for r steps we shall have

$$H_\alpha^{(r)} = r\, H_\alpha^{(1)} \qquad \alpha > 0, \ \alpha \neq 1.$$

for the various probabilities for r steps from any state to start with can be had as the various terms of $(p_1+ p_2+ p_3)^r$.

When $[p_{ij}]$ is of order n, proceeding in the same manner as in (1) we shall have

$$H_\alpha^{(1)} = (1/1-\alpha) \log (p_1^\alpha + \ldots\ldots + p_n^\alpha) \quad \alpha > 0, \alpha \neq 1$$

and arguing the same way as above we shall have for this case

$$H_\alpha^{(r)} = (1/1-\alpha) \log (p_1^\alpha + \ldots + p_n^\alpha)^r = r H_\alpha^{(1)}, \ \alpha > 0, \ \alpha \neq 1 \quad (13)$$

In the limit, as $\alpha \to 1$, (13) tends to the known result $H_1^{(r)} = r\, H_1^{(1)}$. Hence combining this symbolically with (22) we shall have (12).

Q. E. D.

<u>Remark</u> Shannon in his definition of entropy

$$H_1(X) = - \sum_i p(x_i) \log p(x_i)$$

covered the case, when some of the $p(x_i)$ are zero, by laying down the convention that $0 \times \infty = 0$. This is otherwise true in the extended real number system. In the same fashion if we adopt the same convention or consider the evaluations of our α-entropies in the extended real number system, Renyi's α-entropy when some $p(x_i) = 0$ and hence our α-entropies given by (2) and (3) would be valid even if ~~the~~ some of the p_{ij} are zero.

PART - II

1. α-ENTROPY AND ITS EXISTENCE

In section 1.1 α-entropy associated with a stationary source is defined and in section 1.2 its existence is established for a source when the channel through which its contents are to be transmitted has a symmetric noise structure.

1.1 α-Entropy

Consider the set $A^I = \{(\ldots, x_{-1}, x_0, x_1, \ldots)\}$ of all sequences, infinite on both sides, constructed from a set A of the letters of a given alphabet. The subset of all those

sequences of A^I everyone of which has specified letters at certain specified positions is known as a set or briefly a cylinder. If F_A be the smallest Borel field containing all such cylinders then we know that a probability measure determining the probabilities of all such cylinders will uniquely determine the probability for every member of F_A. Let μ be such a probability measure. Further a sequence $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$ is called stationary if its probability remains invariant under a shift operator T (say) i.e. if $\mu(T x) = \mu(x)$.

So, following Khinchin (1957), a discrete stationary source is defined as ~~a doubly stationary source is defined as~~ a doubly infinite stationary sequence of random variables $\ldots X_{-1}, X_0, X_1, \ldots$ taking values on a finite set A called the alphabet of the source.

Suppose A has a letters and consider all n terms sequences; they will obviously be a^n in number. Every such sequence C(say) of the whole lot of a^n in number is a cylinder set in the space A^I consisting of all $x \in A^I$ for which the $x_t, \ldots, x_{t+n-1}$ assume the fixed values characterizing the given sequence C, and therefore has a definite probability $\mu(C)$. So the set of all such n-term sequences represents a finite probability space of a^n elementary events C with probabilities $\mu(C)$. Hence following the procedure adopted for getting Renyi's α-entropy

$$H_\alpha^n = (1/1-\alpha) \log_2 \sum_C \mu^\alpha(C) \qquad \alpha > 0, \alpha \neq 1 \qquad (1)$$

associated with such a space.

If the source emitting these n-term sequences is stationary then H_α^n in (1) will be independent of the moment the source started working and depends only on n and the nature of the source. So the average amount of α-entropy accrued per symbol from such a source would be H_α^n/n. Since n is arbitrary and there is no limit to its choice, so different values of n will give rise to different values of H_α^n/n and thus we shall get a sequence of H_α^n/n. If this sequence converges, it would be natural to call the limit, to which it converges, as the α-entropy per symbol of this source. Denoting it by H_α we have

$$H_\alpha = \operatorname*{Lt}_{n \to \infty} H_\alpha^n/n \quad , \qquad \alpha > 0, \alpha \neq 1 \qquad (2)$$

The following section will prove that H_α exists for a stationary source when the conditional probability matrix obtained from combining two stationary sources consists of identical rows irrespective of the arrangement of their elements.

2.2 Existence

It is given in the form of the following theorem:

Th. 1 Establish the convergence of H_α^n/n for a stationary source identified with a stationary symmetric channe.

Proof. The space S_{m+n} of sequences of length m+n can be regarded as the product of the space S_m of sequences of length m and the space S_n of sequences of length n. Hence using (47) and (75) of chapter 2, we shall have

$$H_\alpha(S_{m+n}) \leqslant H_\alpha(S_m) + H_\alpha(S_n) \qquad \alpha > 0,\ \alpha \neq 1$$

In the present notation this can be written as

$$H_\alpha^{m+n} \leqslant H_\alpha^{m} + H_\alpha^{n} \tag{3}$$

Using (37) of chapter 2

$$H_\alpha^{n} \leqslant H_\alpha^{m+n} \tag{4}$$

Putting $m = 1$, (4) reduces to

$$H_\alpha^{n} \leqslant H_\alpha^{n+1} \tag{5}$$

and putting $m = n$, (3) reduces to

$$H_\alpha^{2n} \leqslant 2\, H_\alpha^{n} \tag{6}$$

By repetitive use of (5), (6) can be generalised to

$$H_\alpha^{nk} \leqslant n\, H_\alpha^{k} \tag{7}$$

Putting $k = 1$, (7) reduces to

$$H_\alpha^{n} \leqslant n\, H_\alpha^{1}$$

$$\text{or} \quad H_\alpha^{n}/n \leqslant H_\alpha^{1} < \infty \tag{8}$$

(8) and the fact that $H_\alpha^{n}/n \geqslant 0$ establish that H_α^{n}/n is bounded above and below so by Boltzano-Weistrauss theorem the seq. $\{H_\alpha^{n}/n\}$ has both upper and lower limit. Let

$$a = \underset{n \to \infty}{\mathrm{Lt}}\ \mathrm{inf.}\ H_\alpha^{n}/n$$

so to prove the convergence of $\{H_\alpha^n/n\}$ it is sufficient to show that $\operatorname{Lt}_{n\to\infty}$ sup. $H_\alpha^n/n = a$. For an $\epsilon > 0$, we can choose an integer r such that

$$|H_\alpha^r/r - a| < \epsilon$$

In particular this can be written as

$$H_\alpha^r / r < a + \epsilon \tag{9}$$

For an n we can select an integer $k > 0$ such that

$$(k-1)\, r < n \leq k\, r \tag{10}$$

From (10) and (5) we have

$$H_\alpha^n \leq H_\alpha^{kr}$$

This combined with (7) would give

$$H_\alpha^n \leq k\, H_\alpha^r$$

Or

$$H_\alpha^n/n \leq kr/n \quad H_\alpha^r/r \tag{11}$$

Using (9) and (10), (11) reduces to

$$H_\alpha^n/n < (k/k-1)(a + \epsilon) \tag{12}$$

Hence for sufficiently large n (and hence for appropriately large k) (12) reduces to

$$a - \epsilon < H_\alpha^n/n < a + 2\epsilon$$

and from this for arbitrarily small ϵ we shall have

$$\underset{n \to \infty}{\mathrm{Lt}} \ \mathrm{Sup.} \ H_{\alpha}^{n} / n = a$$

Q.E.D.

2. α-INFORMATIONS

Consider again the transmission through a channel of the members of $A^I = \{(\ldots, x_{-1}, x_0, x_1, \ldots)\}$. To every transmitted sequence x there corresponds a received doubly infinite sequence belonging to the set $B^I = \{(\ldots, y_{-1}, y_0, y_1, \ldots)\}$ constructed from the output alphabet B. The probability that in the received sequence $y_n = b$ may be regarded in general to depend on the whole sequence x and is the conditional probability that the sequence $y \in B^I$ of received signals will belong to cylinder $y_n = b$. But to characterise the channel completely we must know not only all such probabilities but also of more complex sets $S \in F_B$ (smallest Borel field containing B^I).

Hence following Khinchin (1957) a channel is completely specified if we know its (i) input alphabet A (ii) output alphabet B and (iii) the set of all probabilities $\vartheta_x(S)$ for any $x \in A^I$ and any $S \in F_B$ and it is symbolized by $[A, \vartheta_x, B]$ and the coupling of the channel $[A, \vartheta_x, B]$ and the source $[A, \mu]$ gives rise to a new source $[C, w]$ termed as the compound source where $C = A \times B$, $C^I = A^I \times B^I$ and $w(S)$, for every $S \in F_C$, is given by

$$w(S) = w(M \times N) = \int_M \vartheta_x(N) \; d\mu(x) \qquad (13)$$

Putting $M = A^I$ and leaving $N \in F_B$ arbitrary, we shall have $w(M \times N)$ as the probability of the joint occurrence of

$x \in A^I$ and $y \in N$; but the former is the source so $w(M \times N)$ may be taken equal to $\eta(N)$ as the probability of occurrence of $N \in F_B$ at the channel output . Thus $\eta(N)$ plays the same role for the space B^I as $\mu(M)$ does for the space A^I. Hence (13) reduces to

$$\eta(N) = w(A^I \times N) = \int_{A^I} \vartheta_x(N)\, d\mu(x)$$

and this gives rise to the source $[B, \eta]$ at the channel output. We also know from Khinchin (1957) that if $[A, \mu]$ and $[A, \vartheta_x, B]$ are stationary then $[C, w]$ and $[B, \eta]$ are also stationary.

In section 1 we have established that every stationary source feeding a stationary symmetric channel has a definite α-entropy. So if $[A, \vartheta_x, B]$ is effected by symmetric noise then $[A, \mu]$ would have definite α-entropy attached to it. Following the same procedure as used for establishing the existence of $\mathrm{Lt}_{n \to \infty} H_\alpha^n / n$ we can prove that there would be definite α-entropy attached to $[B, \eta]$. The validity of (3) for symmetric channels would make the α-entropy associated with $[C, w]$ exist.

We also know that specifying the ''pair sequence'' $x_0, y_0, \dots, x_{n-1}, y_{n-1}$ is equivalent to specifying the pair of sequences $x_0, x_1, \dots, x_{n-1}$ and $y_0, y_1, \dots, y_{n-1}$, so the space of sequences $x_0, y_0, \dots, x_{n-1}, y_{n-1}$ is the product of the spaces of sequences $x_0, \dots, x_{n-1}$ and $y_0, y_1, \dots, y_{n-1}$. In other words the scheme constituting the elementary events of the type $x_0, y_0, \dots, x_{n-1}, y_{n-1}$ is the outcome of a random

association of the scheme constituting elementary events of the type $x_o,\ldots,x_{n-1}$ and the scheme constituting elementary events of tye type $y_o,\ldots,y_{n-1}$. Hence following the procedure adopted for deriving (47) of Chapter 2, we can establish that

$$\overset{n}{H}_{\alpha}(Y/X) \leq \overset{n}{H}_{\alpha}(Y) \qquad \alpha > 0,\ \alpha \neq 1 \qquad (14)$$

and similarly

$$\overset{n}{H}_{\alpha}(X/Y) \leq \overset{n}{H}_{\alpha}(X) \qquad \alpha > 0,\ \alpha \neq 1 \qquad (15)$$

From (14) follows the existence of

$$\underset{n \to \infty}{\mathrm{Lt}} \frac{\overset{n}{H}_{\alpha}(Y/X)}{n} = H_{\alpha}(Y/X) \quad \text{(say)} \quad \alpha > 0,\ \alpha \neq 1.$$

and from (15) follows the existence of

$$\underset{n \to \infty}{\mathrm{Lt}} \frac{\overset{n}{H}_{\alpha}(X/Y)}{n} = H_{\alpha}(X/Y) \quad \text{(say)} \quad \alpha > 0,\ \alpha \neq 1.$$

So based on (14) and (15) there would be two types of α-informations conveyed about X by Y and information about Y by X i.e.

$$\overset{n}{I}_{\alpha}(X;Y) = \overset{n}{H}_{\alpha}(X) - \overset{n}{H}_{\alpha}(X/Y) \qquad \alpha > 0,\ \alpha \neq 1 \qquad (16)$$

and

$$\overset{n}{I_\alpha}(Y;\ X) = \overset{n}{H_\alpha}(Y) - \overset{n}{H_\alpha}(Y/X) \qquad \alpha > 0,\ \alpha \neq 1 \tag{17}$$

Dividing (16) and (17) by n and taking limits, as $n \to \infty$, we shall have α-informations per symbol of type I and type II. Denoting them by $I_\alpha(X;\ Y)$ and $I_\alpha(Y;\ X)$ we have

$$I_\alpha(X;Y) = \underset{n \to \infty}{Lt} \left[\frac{\overset{n}{H_\alpha}(X)}{n} - \frac{\overset{n}{H_\alpha}(X/Y)}{n} \right]$$

$$I_\alpha(Y;\ X) = \underset{n \to \infty}{Lt} \left[\frac{\overset{n}{H_\alpha}(Y)}{n} - \frac{\overset{n}{H_\alpha}(Y/X)}{n} \right]$$

$I_\alpha(X;\ Y)$ and $I_\alpha(Y;\ X)$ may also be termed as rates per symbol (or signal) of α-informations as conveyed about X by Y and as conveyed about Y by X.

REFERENCES

1. Abramson, N. (1963) : Information Theory and Coding, McGraw Hill Book Company Inc., New York.

2. Anderson, T.W. (1958) : An Introduction to Multivariate Statistical Analysis, John Wiley and Sons.Inc.,N.Y.

3. Ash, R.B. (1965) : Information Theory, Interscience Publishers, a division of John Wiley and Sons. Inc., N.Y.

4. Barnard, G.A.(1951): The Theory of Information, J. Roy.Statist.Soc.,Ser.B,Vol.13, pp. 46-64.

5. Beckenbach, E.F. (1942) : Convexity Properties of Generalised Mean Value Function, Ann.Math. Statist.,Vol.13, pp. 88-90.

6. Beckenbach, E.F., Bellmen, R. (1965) : Inequalities, Second Edition, Springer-Verlag, Berlin.

7. Campbell, L.L. (1965) : A coding Theorem and Renyi's Entropy, Information and Control, Vol. 8, No.4, pp. 423-429.

8. Chhabra, P.N.(1968): α-Entropies For Discrete Memoryless Channels, Accepted in Labdev in two parts.

9. Cramer, H.(1958) : Mathematical Methods of Statistics, Princeton University Press, Princeton.

10. Duffin, R.J., Peterson E.L. and Zener, C. (1967). : Geometric Programming; Theory and applications, John Wiley and Sons, Inc., N.Y.

11. Fadeev, D.K.(1957) : Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschema, Abheiten zur Informations theorie I Berlin, Deutscher Verlag der Wissenschaften,pp. 85 - 90.

12. Feinstein, A.(1958): Foundations of Information Theory, McGraw Hill Book Company, Inc., N.Y

13. Gel'fand,I.M. and Laglom, A.M. (1959) : Calculation of the Amount of Information about a Random Function Contained in Another Such Function, Trans. Am.Math.Soc.,Ser.2, Vol. 12, pp. 199-246.

14. Goldman, S. (1953) : Information Theory, Prentice Hall, Inc., Englewood Cliffs, N.J.

15. Hardy, G.H., Littlewood: J.E. and Polya G. (1952) : Inequalities, Second Edition Cambridge University Press, Cambridge.

16. Kapur, J.N. (1967a) : Generalized Entropy of Order α and type β., The Mathematics Seminar IV, 78 - 94.

17. ______________ (1967b) : On the Postulates of Entropy in Information Theory, The Mathematics Seminar IV, pp.95 - 104.

18. ______________ (1968a) : On Some Applications of Dynamic Programming to Information Theory, Proc. Ind.Acad.Sciences, Vol.67, No. 1, pp. 1-11.

19. ______________ (1968b) : Some Aspects of Mathematics of Operations Research, Presidential Address at the fifty-fifth Session of the Indian Science Congress, Varanasi.

20. Khinchin, A.I.(1957): Mathematical Foundations of Information Theory, Dover Publications, New York.

21. Kreyszig, E. (1962) : Advanced Engineering Mathematics, John Wiley and Sons, Inc., New York.

22. Lee, P.M. (1964) : On the Axioms of Information Theory, Ann.Math.Stat., 35, pp. 415-418.

23. Littlewood, J.E. (1944) : Lectures on the Theory of Functions, Oxford University Press, Oxford.

24. McGill, W.J.(1954) : Multivariate Information Transmissio IRE Trans.Inform.Theory, Vol.4, pp. 93 - 111.

25. Morrey, C.B.Jr. (1966) : Multiple Integrals on the Calculus of Variations, Springer-Verlag, Berli

26. Renyi, A. (1961) : On Measures of Entropy and Information, Proc. Fourth Berkeley Symposium on Math. Stat. and Prob., pp. 547-561.

27. Reza, F.M. (1961) : An Introduction to Information Theory, McGraw-Hill Book Company Inc., N.Y.

28. Shannon, C.E.(1948) : A Mathematical Theory of Communication, Bell System Tech. J., Vol. 27, pp. 379-423, 623-656.

29. Shniad, H.(1948) : On the Convexity of Mean Value Functions, Bull.Am. Math.Soc., Vol.54, pp.770-776.

30. Varma, R.S.(1965) : Entropy in Information Theory, Ganita, 16, pp. 1-6.

31. __________(1966a) : Generalisations of Entropy of Order α, J. Math. Sciences, Delhi.

32. __________(1966b) : Mathematics, Biology and Information Theory, Presidential address, 32nd Session of the Indian Mathematical Society, Patiala, 1966.